AISB Convention 2017
University of Bath, April 19-21
http://aisb2017.cs.bath.ac.uk/index.html

"Social Aspects of Cognition: Human and Artificial Life" Symposium

This Symposium (which is the continuation of the past AISB2015 and AISB2016) aims at stimulating a lively discussion on the social dimension of knowledge, behavior and ontology by crossing Philosophy, and AI. We point on the following topics (but not exclusively):

I. Strategies for analyzing the problem of the relationship between language, society and AI. Searle presented an interesting theory of representation based on the mind's capacities to represent objects and to the linguistic capacities to extend the representation to social entities. Brandom introduces compelling notion of representation in social terms and explores the differences between human and artificial mind. Moreover, we would like to focus on the issues of the embodiment and embodied cognition (Clark) and the role of social and bodily dimension in linguistic meaning in AI perspectives (Cangelosi et al., Minski). Emotions play a fundamental epistemological role in the "unspoken dimension".

II. The later philosophy of Wittgenstein, classical pragmatism and contemporary analytic pragmatism provide fruitful conceptions of social practices. Social epistemology stresses on their role in human cognition to motivate the overcome of classical individual epistemology. Several important notions are analyzed (social behavior, social norms, testimony, etc.). McDowell and Davidson mention the role of the social within the process of acknowledgement. There are several examples of the bias between AI and human ability to react on various different problems, such as the problem of translation. Google translation generator is capable to of translatinge from one language to another only to certain extent. When it comes to larger textual corpus, Google ultimately fails to produce meaningful contents. What is it that human possesses and AI does not? Wittgenstein would perhaps call it a form of life, a social dimension. If human cognition is preconditioned by this social dimension, what this social dimension is? How does AI intelligence respond to these social precondition of human knowledge; or how is the absence of social aspect limiting for AI? What is it that human possesses and AI does not? Wittgenstein would perhaps call it a form of life, a social dimension. If human cognition is preconditioned by this social dimension, what this social dimension is? How does AI intelligence respond to these social precondition of human knowledge; or how is the absence of social aspect limiting for AI?
"Can sociological approaches rooted in cybernetics and systems-theory (Luhmann) give us a direction for analyzing, describing and constructing social systems inhabited by humans and machines?"

III. Interactions on social media such as Twitter or comment sections differ from "natural" dialogues in ways which present challenges for theories of discourse, dialogue and argumentation such as RST, SDRT or models influenced by Traum, Cohen & Levesque, Walton & Krabbe and others; as well as philosophical (Brandom, Habermas) semiotic (Eco) and linguistic (Crystal, Spilioti) approaches. On the one hand, participation in online dialogue is typically fluid; interlocutors can join or leave a conversation without formality, it is problematic to assign distinct roles such as "addressee" or "overhearer" or 2nd vs 3rd person, participants may be unknown to each other and have only sparse models of each others' backgrounds and beliefs. On the other, there may be direct access to participants' posting history from which can be extracted a (possibly partial and/or inconsistent) "commitment store" in the sense of Hamblin or Walton & Krabbe.

Understanding Learning Strategy as Conversation

Steve Battle [1]

Abstract. The goal of this research is to understand the learning strategies employed by students using so called Intelligent Tutoring Systems (ITS), with the aim of supporting efforts to improve the effectiveness of this type of educational technology. An experiment is conducted in which we observe and record online interactions between the participant and the ITS, with the goal of discovering learning strategies used by students. Gordon Pask's *Conversation Theory* defines a social constructivist account of learning and teaching, and the hypothesis explored here is whether analysis of the observed data encoded using CT will help to identify behaviours that represent meaningful learning strategies. These strategies are identified with Markov Processes, enabling the use of *Hidden Markov Models* to discover clusters of distinct patterns of behaviour within the observational data.

1 INTRODUCTION

Intelligent Tutoring Systems (ITS) are teaching software designed to take a student through a series of exercises; facilitating *learning-by-doing* [38]. These exercises are carried out in a simulated environment supporting active learning. We investigate the use of an ITS known as SQLTutor that teaches students 'Structured Query Language' for databases by taking them through a series of practical exercises.

Our goal is to model the *conversations* between the learner and the ITS to identify recurring patterns of interaction with specific *learning strategies*. The foundation of this research is the work of Gordon Pask on *Conversation Theory* [25]. This is a *social constructivist* theory of learning [41, 3] in which the role of the teacher, and the conversations between teacher and student, are made explicit. According to Pask [26], "Learning depends on the strategies used by a student in order to direct his [or her] attention" and "at the other extreme, the strategies may be imposed upon the student as *teaching strategies* by a programme, a teaching device, or a training routine." The term *learning strategy* used in this paper acknowledges this dialectic between teaching and learning strategy in the observed data.

We are particularly interested in the way that a learning strategy may switch between different levels of learning. Pask's *Conversation Theory* [27] states that a conversation can exist on many levels [34], with the lowest level engaged directly with a learning environment. In Conversation Theory, learning becomes a process of negotiation with peers and teachers to construct "shared meaning" [36]. We can relate this to Bloom's Taxonomy of Learning Domains [4, 40] in which there are six major categories of *cognitive* learning processes including simple *recall*, *understanding* and *application* of knowledge, upon which higher level processing involving *analysis*, *evaluation* and ultimately *creativity* are built. These higher level process of comprehension require an awareness of context and the place of this knowledge in the overall *curriculum* [15].

For the purposes of this research, the observational data [6, p.82] may be thought of as a stochastic process [9]. A learning strategy can then be identified as a Markov Process [2] enabling the use of *Hidden Markov Models* (HMMs) to "automatically discover which strategies are used in practice." [5] Furthermore, recognising that learners may use different strategies according to the context and tasks they face, HMM *Clustering* is used to identify and cluster different subsequences into a number of distinct learning strategies [37, 39]. It is hoped that future development of this theory may support a better understanding of how students learn [12, p.9].

2 CONVERSATION THEORY

"Language does not transmit information" claims Humberto Maturana [23, p.57], "its functional role is the creation of a cooperative domain of interactions between speakers." Shannon's Information Theory [35] tells only part of the story, quantifying the information capacity of the channel alone, whereas for learning we are interested in the level of order in the system as a whole [33], comprising both the learner and the teacher. The theory of *cognitive constructivism* rejects the simplistic *transmission* model of information in favour of a model pioneered by Piaget [29, p.77] where, "knowledge results from continuous construction."

Building on Piaget, David Kolb states [17, p.49], "Learning is the process whereby knowledge is created through the transformation of experience." Kolb's Experiential Learning Cycle in Figure 1 is a four-stage cycle describing an effective learning experience. It begins i) with a new or revisited *concrete experience*, followed by ii) observation of, and reflection upon the experience leading to iii) new concept formation, which may in turn iv) suggest new hypotheses which can then be tested in the concrete environment.

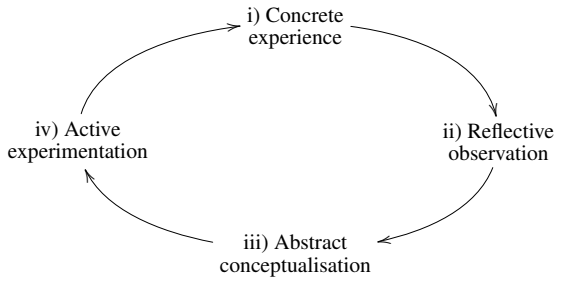


Figure 1: The Kolb Experiential Learning Cycle.

[1] The University of the West of England, Bristol, UK, email: steve.battle@uwe.ac.uk

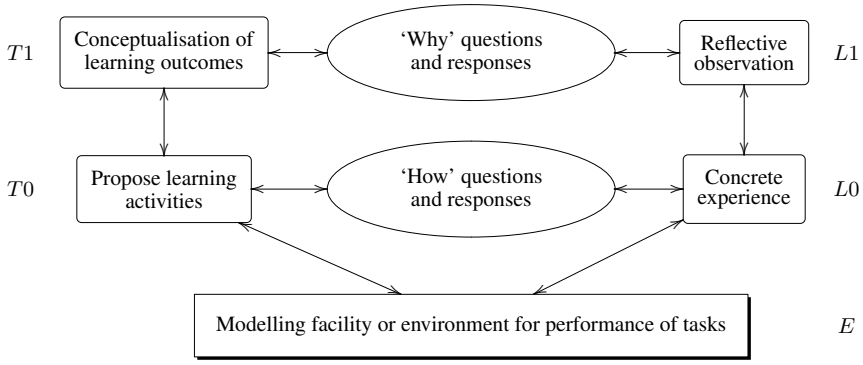


Figure 2: The Skeleton of a Conversation

However, the Kolb cycle is open to criticism [16], the first being that the cycle neglects social learning, particularly the conversation between teacher and learner. The Kolb cycle focuses on the learner as though they were working in isolation. Furthermore, the stages may, in reality, occur out of sequence or not at all. For example, the learner may perform any number of concrete experiments and observations, before later reflecting on the experience as a whole. This would be typical of a practical session where a series of experiments might be conducted en bloc. The reflective stage might be skipped altogether if nothing unexpected occurred that would require the underlying conceptualisation to be modified. Indeed, a learner engaged purely in *surface* learning may simply decide to skip the reflective stage for other reasons; with the reflective and concept formation phases seen as *deep* learning [21, 22, 32].

The theory of social constructivism, developed by Lev Vygotsky [42] rejects the idea of transmission but embraces that of *conversation*. It argues that learning takes place because of a learner's interactions within a group. It supports the idea of reaching understanding by consensus, and it is this idea that finally makes sense of the role of the teacher, and of peers in a learning community. Furthermore, Vygotsky rejected Piaget's notion that a child's *egocentric* language disappears altogether, countering that this private speech is simply a stage in the development of inner speech; an *internal* conversation in all major respects [41].

The cause of Social Constructivism was taken up by British Cybernetician, Gordon Pask, who saw intelligence as an emergent property of *conversation*. From the mid-1950s onward, Pask developed a wide range of teaching machines, the first of which, the Solartron Adaptive Keyboard Instructor (SAKI), led eventually to the development of the world-famous Mavis Beacon typing trainer [30]. His research focused on understanding conversations between teacher and learner, learning peers, a teaching machine interacting with a learner, or in the extreme as an internal dialogue in the mind of an autodidact. Because he is interested in adaptation at a psychological or conceptual level, Pask's *Conversation Theory* [25] is very applicable to these learning conversations. We will use Pask's theory as a framework for understanding the conversation between the learner and an Intelligent Tutoring System in the role of teacher.

Sheila Harri-Augstein and Laurie Thomas, for The Centre for the Study of Human Learning (CSHL) at Brunel University, describe a *learning conversation* [14] as means of recalling experiential memories into awareness, thus being a key enabler for reflection. Moreover, they see learning itself as a skill which can be learned. However, unlike Kolb's Learning Styles and *Experiential Learning cycle*, they view this as a set of parallel conversations occurring at different levels, with task-level conversations ranging over practical activities, and higher level conversations focusing on the learning process itself. We can relate this stratification to Bloom's taxonomy with the task-focused conversation spanning simple recall, through to comprehension involving an appreciation of the wider context, and the place of this knowledge in the overall teaching curriculum. They also observed that traditional teaching practices based on an expert *communicating* their knowledge, may inadvertently reduce a learner's capacity for self-organised learning. This result emphasises the need for good *conversation design* [20] in the development of Intelligent Tutoring Systems.

In his *Conversation Theory*, Gordon Pask develops a similarly stratified model, but one in which the participants of the conversations are explicitly identified. Figure 2 illustrates what Pask calls the 'Skeleton of a Conversation'. The rounded rectangular boxes represent 'Psychological Individuals', or what Pask calls *P-Individuals*. Pask is not particularly concerned with specific learning conversations nor the number of levels, so with Scott [34] we adopt a model that includes individuals that can be loosely identified with Kolb's stages. Both 'Concrete experience' and 'Propose learning activities' deal with the problem of operational learning and involve reasoning directly about action. The 'Conceptualisation of learning outcomes' and 'Reflective observation' work at a higher level, and are concerned with overall comprehension; and are placed correspondingly higher up in the diagram.

The ellipses represent verbal exchanges, or *conversational domains* that psychological individuals are able to engage in. Only individuals at the same level, connected horizontally, interact with each other in this way. The high-level conversations deal with questions about 'Why' the world works the way it does, while the operational levels deal in questions about 'How' the world works; each level having it's own distinctive language.

The vertical connections represent non-verbal, causal connections with feedback. The behaviour of the lower-level process can be reflected upon and selectively modified in the light of some kind of failure; for example, when some action fails to produce the predicted outcome. To ground the operational conversation, individuals at the lowest level level interact with a real or simulated environment; indicated by the heavier box at the bottom of Figure 2. Interaction with this environment is similarly understood as a causal or physical interaction.

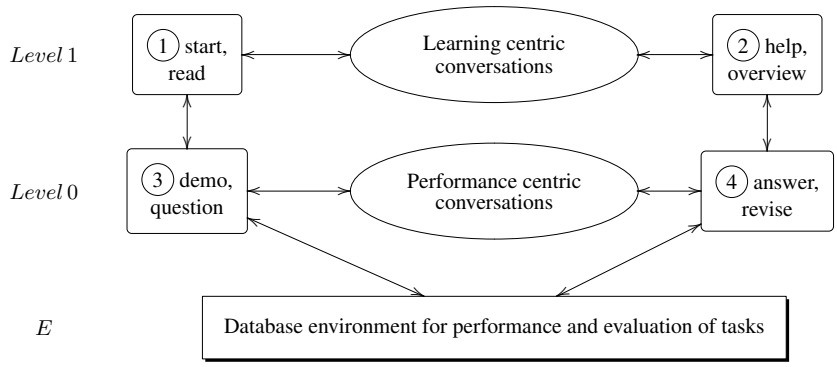


Figure 3: Conversational framework used to encode the data for training the Hidden Markov Models. Each conversational act belongs to one of four classes discriminated by the level of conversation and teaching vs. learning strategy axes.

The P-Individuals in Figure 2 are labelled $T0$, $T1$, $L0$, $L1$ to reflect the most obvious configuration in which the left hand side represents the 'Teacher' ($T0$ & $T1$) and the right-hand side, a 'Learner' ($L0$ & $L1$). Where Kolb describes a solitary learner, this configuration allows us to treat $T1$ as being concerned with the Intended Learning Outcomes whereas $L1$ represents the learners own internally constructed and tacit model of their experience. The Teacher may observe the learner's progress and at $T0$, ask the learner leading operational questions or suggest further experiments. The learner, $L0$, puts these suggestions into practice through interaction with the learning environment E. In this case the processes $T0$ and $T1$ are grouped together as are $L0$ and $L1$. Pask called these groupings *M-Individuals*, or 'Mechanical Individuals', named in a bold attempt to capture both biological and non-biological embodiments of P-Individuals, in a move that allows us to admit teaching machines into the role of Teacher.

Following Kolb more literally, we may arrange individuals $T0$, $T1$, $L0$, $L1$ within a single group – within the mind of an autodidact or self-learner. The self-learner internalises both the teacher and learner, and learning proceeds through an internal dialogue between them. Execution of the cycle would have $L0$ engaging with the environment E; a performance error at this point would be fed up a level to $L1$ for reflection, and then across to $T1$ with the self-learner asking themselves *why* the error occurred. Modifications to the conceptual model may suggest new experiments in the environment E at $T0$ which are then proposed to $L0$, and the cycle repeats.

Conversation Theory also allows us to describe scenarios that are more student-centric, focusing on peer-to-peer collaborative conversations. Collaborative tools are suited to extending the conversation beyond the lecture theatre and have a place in supporting teams of students working together on projects. These are generally conversations at the conceptual level, with $T1$ and $L1$ understood as learning peers; with the model admitting any number of participants into the conversation.

Collaborative learning environments define define a space where learning opportunities are provided in the form of technological artefacts and computer-mediated environments. The so-called flipped classroom [11] focuses on an environment E containing *tangible* artifacts, and *ponderables* introduced by the teacher as facilitator $T0$. Drawing directly on constructivist theories, it focuses on the learning experience of the learner $L0$ and their peers where "the involvement of others can become both catalyst and catharsis in social learning contexts." [13] As before the teacher has a defined set of Learning Outcomes defined at $T1$, and there is an expectation that based on their studies outside the classroom the learner brings Learning Incomes [31] to the table at $L1$.

Pask was also concerned with the content of conversations; not the fine-grained structure of knowledge, but the relationships between topics [27]. He describes just two relationships:- *entailment* and *analogical* relationships. A so-called *entailment mesh* represents the important precursors to a given concept [28] and in the ITS this corresponds to the order in which exercises are presented. The sequence of the questions is set by the conversation designer, each question building on previous answers. This idea can be related to the idea of *threshold concepts* [24] acknowledging the necessity of building concepts hierarchically. The ITS can also present examples and ask the student to apply the same idea elsewhere, reinforcing a learned concept through analogy. The resulting heterarchy of topics is more like a web of concepts that can be approached from any angle, and in this case the order is not critical. In the SQLTutor this concept map is represented very simply as a hypertext, and the conversation is the exploration of this curriculum by the learner.

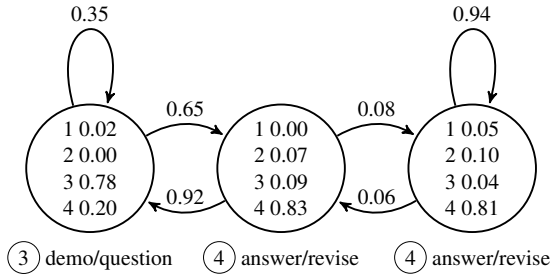
3 EXPERIMENT

The primary data for this research are collected through observation and recording of online interactions between the participant and the SQLTutor running in a standard web-browser. SQLTutor guides the student through a number of exercises using a sequence of readings, demonstrations, and questions. These questions must be answered in the form of SQL statements that are evaluated by the software and checked against a model answer. It offers guidance where the student's answers are incorrect. While the student is guided along a path that presents challenges in a logical order, the student is actually free to move backwards in the sequence to revise content, or to break out of the sequence entirely and visit any point in the curriculum; no page is more than 3 clicks away from any other. The student may also consult online help or refer to the data model underlying the worked examples. This observational sequence is recorded in a dataset that is uniquely associated with each participant.

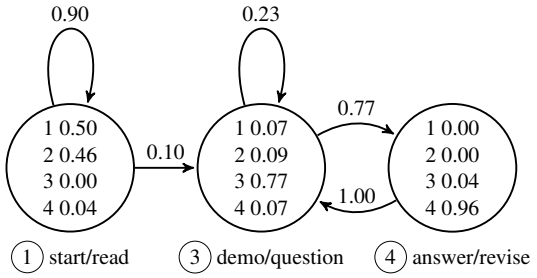
Interactions between the participant and the SQLTutor are captured by a proxy server to which the participant directs their browser for the duration of the session. This provides an electronic record of the Uniform Resource Locators (URLs) visited during the exper-

imental session. A short checklist is presented to the participant before the session, to identify any worksheets they may have previously completed with SQLTutor to estimate their prior knowledge. This information, combined with the observed record is used to determine whether participants are revising existing content or exploring new content.

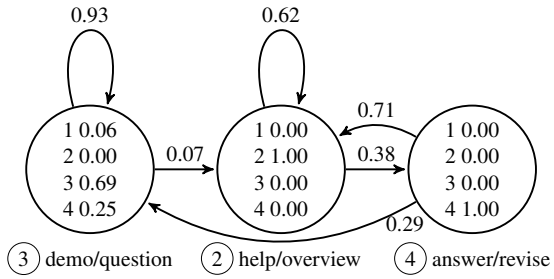
Analysis identifies a number of learning strategies from the observational data. The qualitative intent behind these discovered strategies is then recovered by *triangulating* these observations with follow-up interviews [7, 6]. These semi-structured online interviews, carried out via email [8, p.197], mirror back to those participants relevant excerpts from their observed sessions, asking them what they were trying to achieve. No interpretation was suggested to the participant, leaving them free to provide their own interpretation.



(a) Cluster 1 - *"straightforward Q&A" then "trial and error."*



(b) Cluster 2 - *"seeking help" then "straightforward Q&A".*



(c) Cluster 3 - *"stuck on a question" and "seeking help".*

Figure 5: Three discovered HMM clusters, each one representing a distinct learning strategy. Each state is represented by a node, and transitions between states are shown as directed arcs, labeled with the conditional probability of that transition occurring. In each state a symbol can be emitted, and the probabilities of emitting a symbol (1-4) shown within each state. The symbol with the highest probability is indicated beneath each state along with its conversational acts.

Five subjects participated in the investigation. For analysis these five datasets are merged and segmented at the 'start' points that indicate entry to a new worksheet. Segmenting the data at equivalent points, rather than arbitrarily, allows each HMM to start in the same state. Each visited URL is a page generated by SQLTutor. The recorded URLs are assigned an initial encoding that indicates what kind of page has been visited; the meaning of these *conversational acts* is defined in Table 1. The merged dataset contains 50 sequences, with an average length of 17.28 acts, which are then randomly shuffled to minimise presentation order bias. During training of the HMMs, the shuffled sequences are divided equally into separate training and test sets of 25 sequences each, to reduce over-fitting.

Act	Description
overview	Index listing worksheets, and subjects
start	Worksheet introduction listing contents and related worksheets
read	Topical commentary or summary
help	Help pages and and documentation for the worked example
demo	Demonstration of an SQL command
question	Concrete task proposed by the ITS
answer	Response in the form of an SQL command, and its evaluation
revise	Student revisits an earlier demo, question, or answer page

Table 1: Conversational acts

After Laurillard [19] we re-classify the acts along two dimensions giving us four separate categories as seen in Figure 3. The first (vertical) dimension distinguishes (level 0) task focused or performance centric conversations from higher-level (level 1) conversations which have learning itself as their subject. The learner who lacks the necessary resources to answer a question may explore the range of topics available by returning to a general *overview* page from which they can dip back into and revise specific subject areas. They may also consult general *help* pages accessible from SQLTutor. Performance is grounded in the learning environment, in this case a relational database against which SQL commands can be evaluated, so performance level conversations are those that *demo* a command, ask *questions* & proffer *answers*, or *revise* earlier tasks. Along the second (horizontal) dimension we discriminate between actions initiated by the SQLTutor, such as the *start* of a workshop or an introductory *read*ing, from those initiated by the student learner. This dimension throws into relief the differences between teaching and learning strategies.

This provides us with a reduced set of four encodings, representing four distinct classes of conversational act within Conversation Theory. This merging of conceptually similar acts significantly reduces the size of the search space for Hidden Markov Model clustering.

1. **start/read** (teacher initiated, learning centric)
2. **help/overview** (learner initiated, learning centric)
3. **demo/question** (teacher initiated, performance centric)
4. **answer/revise** (learner initiated, performance centric)

4 HIDDEN MARKOV MODEL CLUSTERING

Hidden Markov Models are used to model potential learning strategies. HMMs are used because we are only able to observe the external actions that appear in a conversation. We do not, of course, have access to the internal 'mental' states of the learner. A HMM is a graphical model where each state is represented by a node, and transitions between states are shown as directed arcs, labeled with the conditional probability of that transition occurring. In each state

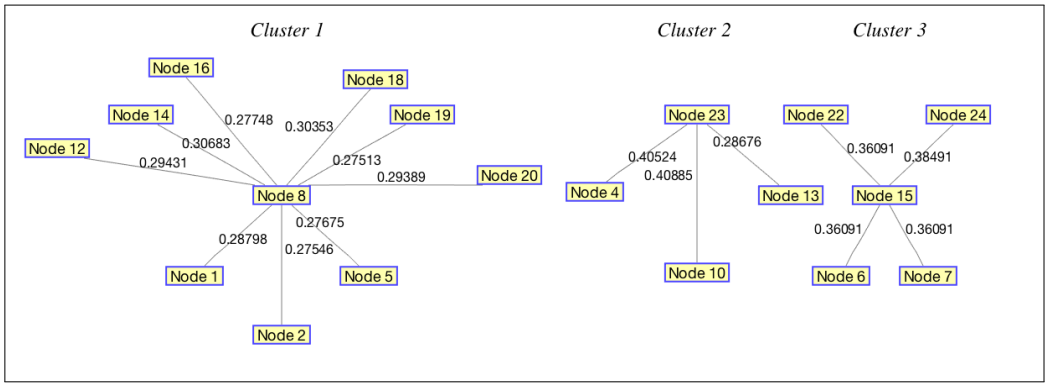


Figure 4: Thresholded Minimum Spanning Tree used to recover clusters. Nodes 1 to 25 represent the segments of observational data from the test set. Another 25 segments comprise the training set. Singleton nodes are not shown.

a symbol can be emitted, here representing one of the four conversational classes of action. The *Markov property* requires that the state at time $t+1$ is *only* dependent on the state at time t; nothing that happened prior to that is relevant. Furthermore, the probability of observing a symbol at time t depends only on the state at time t. If there are any historical dependencies, these must be somehow encapsulated in the current state, in other words the current state is the only 'memory' of the HMM.

Each HMM is defined by a pair $\langle T, E \rangle$ defining a transition matrix T and an emission matrix E. These matrices hold the conditional probability of a given transition, and emission of a symbol, given the current state. The HMMs representing distinct strategies are initially unknown so an HMM clustering algorithm [37] is employed to learn the parameters for a set of HMMs that are, in effect, competing to represent the observed data.

The MATLAB® Statistics and Machine Learning Toolbox[2] is used to train a set of HMMs to recover the (hidden) sequence of states that generated a given set of observed data. The clustering algorithm takes an initially random set of HMMs and partitions the data by finding the best fitting HMM for each datum (a sequence). The number of states and clusters is deliberately kept low (between 1 and 3) to counter over-fitting to the data. The HMM is then trained on that partition (5000 iterations) and the data is re-partitioned and the process repeated for a fixed number of iterations (50). This employs the MATLAB function *hmmtrain* using the Baum-Welch algorithm [1] which is a hill-climbing search that will converge to a local maximum from an initial randomised HMM.

To achieve a global maximum an evidence accumulation technique is used [10] in which co-occurrences of pairs of sequences in the same cluster count as votes towards their eventual association. In addition these co-occurrence counts are weighted by the quality of the solution as evaluated against the test set. After many voting rounds (5000), a consensus global clustering is achieved.

To recover the clusters from the weighted co-occurrence matrix, a minimum-spanning-tree is constructed [18] over the nodes representing the input sequences and the weakest links are cut below a threshold in order to form the clusters shown in Figure 4. The individual nodes in this figure are sequences of observational data from the test set, and it can be seen that there are three clusters, labeled as clusters 1 to 3.

Representative HMMs are then regenerated from the resulting clusters, again using MATLAB's *hmmtrain* function, as shown in Figure 5. Each has 3 states with the transition probabilities indicated on the directed arcs, and the probabilities of emitting a symbol (1-4) (the emission vector) is shown within each state. These symbols correspond to the classes in the framework of Figure 3, and to assist interpretation the symbol with the highest probability is shown beneath each state along with its corresponding conversational acts.

In most cases the HMM training has identified a single high-probability emitted symbol for each state. The only exception to this is the first state in cluster 2 with two approximately equiprobable symbols (1 and 2). This means that the strategy represented by this state does not distinguish between Level 1 learning-centric acts, but does distinguish these from Level-0 performance-centric acts.

4.1 Triangulation by thematic analysis

Given the results of this analysis, the next step was to conduct semi-structured interviews with the participants to obtain their views on the strategies they were using at the time. This enables us to *triangulate* the results of the observational analysis with qualitative comments from each participant, who were asked about representative examples of their own activity falling within different clusters. Not all participants exhibited activity from every cluster. Table 2 summarises the results of these interviews. A simple *thematic analysis* of the comments, generalising across participants, allows us to pull out qualitative features which we can use to describe each cluster.

An NVivo®[3] based analysis enables us to encode the appearance of specific keywords within a comment. These keywords are aggregated to code for certain terms suggested by the discovered learning strategies. These terms may then be combined in ways that reflect the learning strategies more closely. For example, the combination of terms, "seeking help" and "straightforward Q&A" should select for strategies matching cluster 2.

To test the hypothesis that these combinations of terms discriminate between the Hidden Markov Model clusters a contingency table is constructed in Table 3. The rows and columns of this table are nominal values; clusters 1 to 3 are arranged in columns, and each row represents criteria expressed as a term or combination of terms. A row called 'other' captures any unmatched comments, so that all comments are accounted for. The *null* hypothesis is that these criteria and the the clusters are independent. A chi-square test cannot be used

[2] MathWorks http://uk.mathworks.com/products/statistics/

[3] NVivo http://www.qsrinternational.com/nvivo-product

here because the expected values are less than 5, so an *exact contingency table test* is used to calculate the probability of the observed table. The sum of probabilities for this table is $p = 0.008 < 0.05$ (5% significance), so we reject the null hypothesis and conclude that the terms found in thematic analysis are sufficiently discriminatory. This provides additional support for the learning strategies discovered by Hidden Markov Model clustering.

The following points enumerate the terms used to encode the participant's comments, represented as aggregations of keywords.

- **trial & error**
  - "wrong"
  - "trial and error"
  - "see what would happen"
  - "attempts", "attempting"
- **stuck on a question**
  - "stuck"
- **straightforward Q&A**
  - "valid"
  - "trying"
  - "straightforward"
  - "confident"
- **seeking help**
  - "structure of the tables"
  - "look"
  - "help"
  - "familiarise"

Drawing on this thematic analysis we are now in a position to construct an interpretation of the Hidden Markov Models.

Cluster 1 - *"straightforward Q&A" then "trial and error."* in Figure 5(a) begins with a cycle of "straightforward Q&A", of demo/question then answer/revise. There is repetition around 'demo/question' where SQLTutor provides a sequence of demonstrations. However, a difficult question may be followed by a flurry of "trial & error" activity where the same question is answered incorrectly multiple times. Crucially, and for whatever reason, the student is not accessing the help system. According to Sleeman [38] incorrect answers are not always bad, "some floundering can be vitally important. The crucial meta-skill of knowing when one's floundering is useless can only be discovered by trial and error." In other words, this floundering can be seen as a learning strategy in which the learner comes to understand how to apply their knowledge.

Cluster 2 - *"seeking help" then "straightforward Q&A".* in Figure 5(b) is often seen at the beginning of a session where the learner is following the provided reading matter while trying to familiarise themselves with the learning environment. This *serial* activity [26, p.269] around reading is expected as this is the default offered by the available teaching strategy. In the discovered learning strategy this is seen to be followed by a sequence of "confident" question answering where the subject correctly answers all the questions. This combination is most likely to be seen early on when the subject is both unfamiliar with the environment and the questions are simpler.

Cluster 3 - *"stuck on a question" and "seeking help".* in Figure 5(c) contains a question and answer with multiple visits to help pages interposed between them. Even when an (incorrect?) answer has been given, there is high probability of need for further help.

Cluster	Subject	Comment
1	1	"...simply me working through the worksheet normally."
1	1	"...got myself confused using the program, was attempting to recall how to access certain parts."
1	1	"Stuck on a particular question, came down to syntax of SQL statement."
1	2	"to pull information from multiple tables in a single query."
1	3	"likely to be the straightforward questions, or pages that Ive (sic) opened only once."
1	3	"questions that I got wrong multiple times. I used trial and error to get to the answer."
1	4	"The simple SQL statements sections was fairly straight forward."
1	4	"With this cluster I got the arithmetic wrong, ... after a few attempts though I finally realised my mistake and corrected it."
1	5	"I'd forgotten the 'emp' table name, so I just put in the incomplete SQL fragment to see what would happen."
2	2	"I was trying to familiarise myself with SQLTutor and trying out the SQL sandbox."
2	3	"questions that I was confident about. Got all of those on first try."
2	4	"I started off a little confused and began going back and forth between the help sections and the first task."
3	1	"was stuck on a question, so went to look for information within the program."
3	3	"getting familiar with the software and structure of the tables."
3	5	"I just needed to use the DESCRIBE statement for two tables. I already had the 'help' page open..."

Table 2: Triangulation of the results of the observational analysis with qualitative comments from each participant, who were asked about representative examples of their own activity falling within different clusters. Not all participants exhibited activity from every cluster.

Cluster:	1	2	3
trial & error	4	0	0
seeking help ***and*** *straightforward Q&A*	0	2	0
stuck on a question ***and*** *seeking help*	0	0	2
other	5	1	1

Table 3: Contingency table where clusters 1 to 3 are arranged in columns, and each row represents criteria expressed as a term or combination of terms. The discrimination provided by these criteria provides additional support for the learning strategies discovered by Hidden Markov Model clustering.

5 CONCLUSION

This investigation demonstrates the efficacy of using HMM clustering to discover learning strategies in observational records of interactions with an ITS. The benefit of HMMs is that they provide a bottom-up approach to learning strategy discovery, rather than defining them in advance of the investigation. These discovered strategies are seen to recur within sessions and, perhaps more significantly, across multiple students. *Triangulation* enables us to recover student intent via interviews, and verify that these learning strategies have a stable meaning. This confirms the hypothesis that Conversation Theory provides an adequate framework for understanding learning strategies.

ACKNOWLEDGEMENTS

I would like to thank Chris Wallace for the continuing development of SQLTutor[4], used extensively in our teaching of Structured Query Language.

REFERENCES

[1] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, 'A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains', *Annals of Mathematical Statistics*, **41**(1), 164–171, (1970).

[2] Leonard E. Baum and Ted Petrie, 'Statistical inference for probabilistic functions of finite state markov chains', *Ann. Math. Statist.*, **37**(6), 1554–1563, (12 1966).

[3] P.L. Berger and T. Luckmann, *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*, Open Road Media, 2011.

[4] B.S. Bloom, *Taxonomy of Educational Objectives: The Classification of Educational Goals*, number v. 1 in Taxonomy of Educational Objectives: The Classification of Educational Goals, D. McKay, 1956.

[5] Kristy Elizabeth Boyer, Eun Young Ha, Michael D. Wallis, Robert Phillips, Mladen A. Vouk, and James C. Lester, 'Discovering tutorial dialogue strategies with hidden markov models', in *Proceedings of the 2009 Conference on Artificial Intelligence in Education: Building Learning Systems That Care: From Knowledge Representation to Affective Modelling*, pp. 141–148. IOS Press, (2009).

[6] E. Cleaver, M. Lintern, and M. McLinden, *Teaching and Learning in Higher Education: Disciplinary Approaches to Educational Enquiry*, SAGE Publications, 2014.

[7] Louis Cohen, Lawrence Manion, and Keith R. B. Morrison, *Research methods in education (5th edition)*, Routledge, Oxon, UK, 2000.

[8] M. Denscombe, *The Good Research Guide: For Small-scale Social Research Projects*, Open UP study skills, McGraw-Hill Education, 2014.

[9] J.L. Doob, *Stochastic Processes*, Wiley Publications in Statistics, John Wiley & Sons, 1953.

[10] Ana L. N. Fred and Anil K. Jain, 'Combining multiple clusterings using evidence accumulation', *IEEE Trans. Pattern Anal. Mach. Intell.*, **27**(6), 835–850, (June 2005).

[11] P. Freire, *Pedagogy of the Oppressed*, Continuum, 1970.

[12] H. Fry, S. Ketteridge, and S. Marshall, *A Handbook for Teaching and Learning in Higher Education: Enhancing Academic Practice*, Taylor & Francis, 2003.

[13] Susan Greener, 'Flipped or blended? what?s the difference and does it make a difference to learning in he?', in *Proceedings of the 10th International Conference on e-learning (ICEL2015)*, (June 2015).

[14] S. Harri-Augstein and L.F. Thomas, *Learning Conversations: The Self-Organised Learning Way to Personal and Organisational Growth*, Routledge, 1991.

[15] Janet Holland and Marcus Childress, 'conversation theory conceptualized in e-learning environments', in *Handbook of Conversational Design for Instructional Applications*, IGI Global, (2008).

[16] Peter Jarvis, 'Learning to be a person in society', in *Contemporary Theories of Learning: Learning Theorists ... In Their Own Words*, ed., K. Illeris, Taylor & Francis, (2009).

[17] D. A. Kolb, *Experiential Learning: Experience as the Source of Learning and Development*, Prentice-Hall, Englewood Cliffs, New Jersey, 1984.

[18] Joseph B. Kruskal, 'On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem', *Proceedings of the American Mathematical Society*, **7**(1), 48–50, (February 1956).

[19] D. Laurillard, *Rethinking University Teaching: A Conversational Framework for the Effective Use of Learning Technologies*, Rethinking University Teaching: A Conversational Framework for the Effective Use of Learning Technologies, Routledge/Falmer, 2002.

[20] Rocci Luppicini, 'Introducing conversation design', in *Handbook of Conversational Design for Instructional Applications*, IGI Global, (2008).

[21] F. Marton and R. Säljö, 'On qualitative differences in learning: I – outcome and process', *British Journal of Educational Psychology*, **46**(1), 4–11, (1976).

[22] F. Marton and R. Säljö, 'On qualitative differences in learning: II – outcome as a function of the learner's conception of the task', *British Journal of Educational Psychology*, **46**(1), 115–127, (1976).

[23] Humberto R. Maturana and Francisco J. Varela, *Autopoiesis: The Organization of the Living*, Springer Science & Business Media, 1993.

[24] J. Meyer and R. Land, *Overcoming Barriers to Student Understanding: Threshold Concepts and Troublesome Knowledge*, Taylor & Francis, 2006.

[25] G. Pask, *Conversation, Cognition and Learning: A Cybernetic Theory and Methodology*, Elsevier, New York, 1975.

[26] G. Pask, *The Cybernetics of Human Learning and Performance: A Guide to Theory and Research*, Hutchinson Educational, Hutchinson Educational, 1975.

[27] G. Pask, *Conversation Theory: Applications in Education and Epistemology*, Elsevier, 1976.

[28] G. Pask and B.C.E. Scott, 'Caste: A system for exhibiting learning strategies and regulating uncertainties', *International Journal of Man-Machine Studies*, **5**(1), 17 – 52, (1973).

[29] J. Piaget, *Genetic epistemology*, Columbia University Press, New York, 1970.

[30] Andrew Pickering, *The Cybernetic Brain: Sketches of Another Future*, The University of Chicago Press, Chicago, 2009.

[31] P. Race, *The Lecturer's Toolkit: A practical guide to assessment, learning and teaching*, Taylor & Francis, 2014.

[32] P Ramsden, *Learning to teach in higher education*, Routledge, 2nd edn., 2003.

[33] Erwin Schrödinger, *What is Life?*, Cambridge University Press, Cambridge, 1944.

[34] Bernard Scott, 'Conversation theory: A constructivist, dialogical approach to educational technology', *Cybernetics and Human Knowing*, **8**(4), 25–46, (2001).

[35] Claude E. Shannon, 'A Mathematical Theory of Communication', *The Bell System Technical Journal*, **27**(3), 379–423, (1948).

[36] Mike Sharples, 'Learning as conversation: Transforming education in the mobile age', in *Seeing, Understanding, Learning in the Mobile Age*, pp. 28–30, (2005).

[37] Benjamin Shih, Kenneth R Koedinger, and Richard Scheines, 'Discovery of student strategies using hidden markov model clustering', *International Conference on Educational Data Mining*, (2010).

[38] D. Sleeman and J.S. Brown, *Intelligent tutoring systems*, Computers and people series, Academic Press, 1982.

[39] Caitlin Tenison and Christopher J. MacLellan, 'Modeling strategy use in an intelligent tutoring system: Implications for strategic flexibility', in *Intelligent Tutoring Systems: 12th International Conference, ITS 2014, Honolulu, HI, USA, June 5-9, 2014. Proceedings*, eds., Stefan Trausan-Matu, Kristy Elizabeth Boyer, Martha Crosby, and Kitty Panourgia, 466–475, Springer International Publishing, (2014).

[40] Kenneth E. Vogler, 'Improve your verbal questioning', *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, **79**(2), 98–103, (2005).

[41] Lev Vygotsky, *Thought and Language*, MIT Press, Cambridge, MA, 1962.

[42] Lev Vygotsky, *Mind in Society*, Harvard University Press, 1978.

[4] 'The Wallace Line' http://thewallaceline.blogspot.co.uk/2009/02/data-modelling-tutor.html

Does the brain encode information?

Jakub Mácha[1]

Abstract. Our common sense intuition says that when remembering something we store a piece of information in our memory, that is, in our brain. We can go further by claiming that the brain computes a program by processing information. The computational theory of the mind treats minds as information processing systems. These claims are the main tenets of contemporary cognitive science and neuroscience. Drawing on Searle's famous Wall argument, I argue, counter to these claims, that the brain cannot carry information by any reasonable definition of information (Shannon, semantic, algorithmic, quantum). Possessing information cannot be an *intrinsic* feature of any material object, including the brain. If the brain cannot store any information, it cannot compute any function or run any program, for non-trivial functions and programs presuppose the input information. Hence, the analogy between hardware/software and brain/mind is flawed in this respect.

1 INTRODUCTION

There has been always a human desire to explain the mind with the help of an analogy to the most developed technology of the time, and a desire to build a machine that would be (like) a human mind. Once it was a wax tablet, then it was a hydraulic system, a mechanic system, the steam engine, the phonograph, a camera, a hologram. And since the middle of the 20th century, it has been the computer.[2]

The analogy between the human mind and the computer is appealing. Some of its parts are clearly correct and worth further research. The central tenet of this analogy is that the human brain (or body) is analogous to the hardware and the human mind is analogous to software. There is some truth in this. For instance, there is the fact that the causal mechanical processes in the hardware, which cause programs to run, is analogous to the fact that certain (biological, molecular, causal) processes in the brain give rise to the appearance of the human mind. To put this in negative terms, not having software without hardware is analogous to no mind without a brain. This relationship is nowadays called *supervenience*: any change in a mental state implies a change in a physical state. What kind of processes and on which level are these responsible for the appearance of the human mind is one of the big questions facing cognitive science and philosophy today.

The core of the analogy is: Since computers process information, i.e. compute, brains must be able to compute as well. The analogy, thus, depends on the notions of computer and computation. In computer science, computers are usually modelled as Turing machines, the notion of computation remains somehow vague and ambiguous (cf. Searle's "standard textbook definition" of computation below). I shall focus on one component of the Turing machine: the *tape* which is the machine's memory. In Turing's words: "The machine is supplied with a 'tape' (the analogue of paper) running through it, and divided into sections (called 'squares') each capable of bearing a 'symbol.'" [2] In our analogy, the tape is a model of human memory. Here is an outline of the first part of my argument: If brains are computers, they must be able to process information. If they are able to process information, they must be able to store and retrieve it in their memory. In short, the brain must be able to encode information. Before we focus on the notion of information, which is the main topic of this paper, let us discuss a stronger instance of our analogy, which is the Computational Theory of Mind (CTM).

Although there are different versions of CTM, its core can be summed up as follows: the brain is *essentially* a computer (hardware) and the mind is *essentially* a program (software). Mental states and processes are essential computational states and processes. So, for instance, Jerry Fodor says: "Insofar as we think of mental processes as computational (hence as formal operations defined on representations) it will be natural to take the mind to be, *inter alia*, a kind of computer. That is, we will think of the mind as carrying out whatever symbol manipulations are constitutive of the hypothesized computational processes. […] we may thus construe mental operations as pretty directly analogous to those of a Turing machine. There is, for example, a working memory (corresponding to a tape)" [3, p.65]. David Chalmers claims "that computational descriptions of physical systems serve to specify their causal organization (or causal topology), and that cognitive processes and properties are organizational invariants that are determined by the abstract causal organization of the system." [4, p.3]

John Searle, in a series of thought experiments and arguments, argued that the brain-computer analogy and especially CTM are misguided. He came out with the following structure of his argument (which captures his Chinese room argument, although not in the form of a thought experiment):

- Axiom 1. Computer programs are formal (syntactic).
- Axiom 2. Human minds have mental contents (semantics).
- Axiom 3. Syntax by itself is neither constitutive of nor sufficient for semantic content.
- Conclusion 1. Programs are neither constitutive of nor sufficient for minds. [5, p.27]

By this argument, Searle maintained that a computer can produce a perfect simulation of the human mind's cognitive capacities like understanding a language without having these capacities. I take Searle's Chinese room argument to show, *inter alia*, that semantics cannot be ascribed to an agent on the basis of observing its behavior (e.g. on the basis of the Turing test).

[1] Department of Philosophy, Masaryk University, Arna Nováka 1, 602 00 Brno, Czech Republic, email: macha@mail.muni.cz.

[2] See [1] as an exemplary article: "The Memory Code" with the subtitle: "Researchers are closing in on the rules that the brain uses to lay down memories. Discovery of this memory code could lead to the design of smarter computers and robots and even to new ways to peer into the human mind".

As we have seen for Fodor, computational states are defined on representations, i.e. semantically. There are, however, versions of CTM that do not rest on semantics. That is, it doesn't matter what computational states represent or what they refer to. Chalmers clearly endorses this view: "Semantic notions such as that of representation played no role in that account." [4, p.6] The Chinese room argument therefore does no harm for such a version of CTM.
What all versions of CTM have in common is that it is essential for the computation, and ergo for the mind, that the computer manipulates symbols, i.e. that computers are syntactical machines. Hence, Turing, Fodor and Chalmers, all accept Searle's Axiom 1.[3]
In 1992, Searle [7] published an even more radical argument: Any material object can compute any program; for instance, the wall behind me can be *interpreted* as computing the Wordstar program (a word processing program from the 1980s). If this were so, the fact that an object computes a program would be trivial. This argument is sometimes called a triviality argument against functionalism or computationalism about the mind.[4] Why can Searle claim that a wall is implementing the Wordstar program? He presupposes, what he calls the standard textbook definition of computation, which is equivalent to Axiom 1 above: that computation is defined syntactically as a manipulation of symbols. Now, given a syntactical description of an object, we can say "there is some pattern of molecule movements which is isomorphic with the formal structure of Wordstar".
One may wonder, however, what the formal structure of Wordstar actually is. Is it a particular run of the program when it is editing a particular text? I think we can assume that there exists such a description of the wall that is isomorphic to this instance of Wordstar. Or there is a single description of the wall that is isomorphic with the formal structure of any run of Wordstar? Not the actual run of the program, but also every possible run. Chalmers [9] criticized such triviality arguments (especially Putnam's version of them) against CTM that they cannot accommodate possible, but non-actual (counterfactual) situations. Definitely, there does exist a description of the wall where it is isomorphic with a particular run of Wordstar. However, given Chalmers' argument, there is no single description of the wall that would cover every possible run of Wordstar. There is, of course, a single description of any desktop computer that covers every run of Wordstar. Or in today's terms, there is a natural description of a PC so that it is isomorphic to every possible run of MS Word 2016. This amounts to the almost trivial fact that MS Word 2016 can be run on every contemporary PC, but not on every wall.
Did, then, Chalmers provide a knock-down counter-argument against triviality arguments? In my view he didn't, because the notion of computation is not the main focus of Searle's argument. I think the central claims of the argument are: "Syntax […] is not intrinsic to physics. […] Syntax and symbols are observer relative." [10, p.35]
In short, the Chinese room argument says that semantics is observer relative; and Searle's Wall argument says that syntax is observer relative as well. Both aren't intrinsic to a physical system like a brain. Because the notion of computation involves either syntax alone or both syntax and semantics, computation cannot capture the intrinsic features of a brain.
I'm going to argue that we can arrive at these claims without invoking the notion of computation or implementation. I will employ the notion of information instead. As I have argued above, computation is in fact information processing. The notion of information is thus a part of the notion of computation in the following sense: If information is not intrinsic to a system, then computation is neither. The advantages of this approach are that information doesn't involve any temporal or modal dimension that is liable to Chalmers' argument (i.e., that storing or encoding information means encoding any possible piece of information at any time, given physical limits) and that the notion of information is clearly defined. I'm going to consider every relevant notion of information in order to show that they involve syntax or semantics.
There are many conceptions of information. Primarily, we have to focus on those conceptions that are relevant in computer science. These are, no doubt, Shannon information, algorithmic (or Kolmogorov) information and more recently, quantum information. Moreover, I will consider semantic accounts of information as well.

2 SHANNON INFORMATION

Shannon [11] defines the information of a message or symbol S as $-\log_2 p_S$ where p_S is the probability of the occurrence of S. If the probability of the occurrence of S is 0.5, i.e. that it is equally probable that S occurs and that S does not occur, then S generates information of 1 bit ($= -\log_2 0.5$).
First, we have to make clear what counts as a message or symbol. It can be a letter, a binary code, a continuous function, a vector of several functions or any combination of these [11]. In fact, any object can generate information. Let us take for example a toaster that can be in two states: on and off. Let the probability of being on be p and the probability of being off be 1–p. The average amount of information the toaster generates is then $-p.\log_2 p - (1-p).\log_2(1-p)$. If a toaster can generate information, then a neuron can do so as well. Let us assume that a neuron can fire an action potential, and that the probability of its firing is p. Then the neuron can generate information of $-\log_2 p$.
Two interrelated points are our focus now:
(1) Every phenomenon can be interpreted as a symbol and as a bearer of information. I'll call this fact a *syntactic* point.[5]
(2) It does not matter what this symbol means or refers to. This is a *semantic* point.
Let us now focus on the syntactic point. Every phenomenon (situation, state of affairs, event) can be interpreted as a symbol, i.e.

[3] Objection (by J. Haugeland [6]): Computers run programs which are sequences of instructions. They are able to interpret these programs according to an encoding. They do something more than just manipulating symbols. Hence, computers have a syntax as well as semantics. My reply: When a computer runs a program, it gets the program on its input. In abstract terms, a universal Turing machine gets a program on its input tape. There must be an encoding of this input data, i.e. information. Then, however, the computer or the universal Turing only manipulates these symbols. Computational semantics can be, thus, accounted for in syntactic terms.
[4] Putnam [8] came up with a similar and more detailed argument. Nothing in my account rests however on Putnam's version of the argument.
[5] Cf. Dretske [12]: "Any situation may be taken, in isolation, as a generator of information." Or Searle [7]: "The ascription of syntactical properties is always relative to an agent or observer who treats certain physical phenomena as syntactical."

as a bearer of information. This means that there must be an external agent that decides which phenomena are taken or recognized as symbols. "S" may by a symbol, or a random shape on the screen. Lower case "s" may be the same symbol or another symbol. We can consider the toaster as capable of being in two states: on and off. The toaster can encode any number as its temperature. It can be a natural number or even a real number (up to a certain accuracy of measurement). Finally, we can say that a neuron is either firing (emitting its action potential) or not firing. Then a neuron would encode one bit of information (like the toaster capable of being either on or off). There are, however, many ways how to interpret the state of a neuron, i.e. there are many neural encoding schemes.[6] We can look at the duration, the amplitude, or the shape of actions potentials. Or, we can consider the frequency of neuronal firings (spike firings). More importantly, we can, instead of a single neuron, take blocks of neurons as the basic unit of information encoding.

As I have argued above, every computation presupposes an information encoding. In fact, even a stronger claim can be put forward here: that different information encodings lead to different computations on the same physical states. Here is an example of such a case taken from Shagrir [14]:

> Consider a physical system **P** that works as follows: It emits 5-10 volts if it receives voltages greater than 5 from each of the two input channels, and 0-5 volts otherwise. Assigning '0' to emission/reception of 0-5 volts and '1' to emission/reception of 5-10 volts, the physical gate implements the logical *AND-gate*: '0','0' → '0'; '0','1' → '0'; '1','0' → '0'; '1','1' → '1'.
>
> Let us suppose it turns out that flip detectors of **P** are actually tri-stable. Imagine, for example, that **P** emits 5-10 volts if it receives voltages greater than 5 from each of the two input channels; 0-2.5 volts if it receives under 2.5 volts from each input channel; and 2.5-5 volts otherwise. Let us now assign the symbol '0' to emission/reception of under 2.5 volts and '1' to emission/reception of 2.5-10 volts. Under this assignment, **P** is now implementing the *OR-gate*: '0','0' → '0'; '0','1' → '1'; '1','0' → '1'; '1','1' → '1'.

Two different encodings are considered in this example. On the first encoding, the emission/reception of 0-5 volts encodes '0' and the emission/reception of 5-10 volts encodes '1'. On the second encoding, the emission/reception of 0-2.5 volts encodes '0' and the emission/reception of 2.5-10 volts encodes '1'. Let us assume that voltages are intrinsic properties of the physical system (they don't need to be, see below). These encoding are, however, *assigned* by an external agent. Thus, they are not intrinsic to the physical system.[7] A mere choice of the encoding can determine what function is computed by the system, whereas the underlying physical states are the same in both cases.

Shagrir concludes from his example that "the very same physical system **P** simultaneously implements two distinct logical gates". But why simultaneously? The two encodings are mutually exclusive. The system implements either the *AND-gate*, or the *OR-gate*, depending on the choice the encoding. Blackmon [16] proposes to treat the implementation of a computation as "a three-place relation over causal structures of physical things, programs, and mappings". This mapping which maps causally related physical states to symbols is our information encoding. Given this terminological clarification, no physical system can simultaneously implement different computations.

All these considerations lead us to the following: Does information encoding depend on the encoding agent, i.e. is encoding observer-relative? Advocates of CTM have to come up with an account of information that is independent of the encoding agent. It is clear now, however, that Shannon-information is not. This leads to the following conclusion: If one is interpreting a physical state (S, a toaster, a neuron) as a bearer of (Shannon-)information, the agent has to come up with an encoding, i.e. a method of recognizing physical states as symbols.

This is, however, precisely Searle's point: "For any program and for any sufficiently complex object, there is some description of the object under which it is implementing the program." [6, p.208] Now amending this point to the present case: For any information content and for any sufficiently complex object, there is some encoding under which the object encodes this information content.

3 SEMANTIC INFORMATION

The semantic point has sometimes been seen as the reason why Shannon's theory has been accused of neglecting the semantic aspect of information [17, p.1]. It does not matter whether symbol S means "It is a day" or "The number of my hairs is even". And it does not matter whether the fact that the toaster is on means that it is heating or that the sky is blue. The firing of a neuron may mean that this neuron has been stimulated (for example, by another neuron or, in the causal chain, by an object in the visual field). The connection between a symbol and what it means or refers to, can be natural or conventional. A natural connection is a law-like connection between two states (without any intermediate agent), the most common being a causal regularity (but other regularities are also conceivable, e.g. visual resemblance). A conventional connection between two states is a semantic function that mediates between these states and is set up from outside by an agent (or agents). There is a natural connection, for example, between the number of growth rings on a stump and the tree's age (this is an example of environmental information). There are also connections that are partly natural and partly conventional, e.g. visual resemblance may depend on a perspective which presupposes an external agent.

There are various accounts of *semantic* information, e.g. Bar-Hillel and Carnap [17], Dretske [12], Floridi [18], which compute the amount of information according to the probability of obtaining situations, states of affair, events, possible worlds. In contrast to these accounts, what affects the amount of (Shannon-)information are the relative probabilities of the occurrence of a symbol. We may learn about these probabilities by considering what these symbols mean (e.g. we can research the relative probabilities of letters, words, expressions in a language). The semantic accounts of information are, however, not relevant in computer science. When we say that a hard disk has the capacity of 80 GB, all that matters are the probabilities of its states (which are all equal in

[6] See [13] for an overview.

[7] This point is indicated in Dennett's 1969 dissertation: "The electrical characteristics of an impulse sequence, or the molecular characteristics of a nerve fibre could not independently determine what the impulses *mean* [i.e. the semantic point], or what *message* [i.e. the syntactic point] the nerve fibre carries" [15].

this case). It doesn't matter what is stored on the hard disk. Whatever content is stored there, its capacity remains unchanged. In sum, computer science works neither with environmental semantic information, nor with conventional semantic information.

4 ALGORITHMIC AND QUANTUM INFORMATION

There are two other accounts of information that are relevant in computer science: algorithmic information and quantum information. I am now going to argue that neither of these accounts can do without encoding.

The motivation for the algorithmic information theory (Solomonoff [19], Kolmogorov [20], Chaitin [21]) was the problem of assigning a priori probabilities to symbols (in Shannon) or to situations (in Bar-Hillel and Carnap). The problem is that there is no universal method of assigning these probabilities. The main idea is as follows: The *information content* of a string x is the length of the shortest program that gives x on its output on a given reference universal Turing machine. Formally: $\mathrm{AI}(x) = \min_p\{l(p): U(p) = x\}$, where $l(p)$ gives the length of p in bits, U is a universal Turing machine and $U(p) = x$ means that program p, when run on U outputs x. Intuitively, this definition gives us the length of the shortest description of a given string, i.e. the length of the best compression of it.[8]

There might be the hope that algorithmic information would overcome the problem with arbitrary encodings which fuels the core of Searle's Wall argument. Algorithmic information strives after the best encoding of a given string. Instead of choosing an arbitrary encoding, algorithmic information takes the best one. This move allows us to overcome the arbitrariness of the encoding of semantic information. In other words, it allows us to cope with the arbitrariness of encoding within the semantic point (2) above. The present argument is, however, based on the arbitrariness of the encoding of symbols (the syntactic point (1) above). The algorithmic information is the amount of information of a *string*. It is left open here what actually counts as a string, however. Hence, there must be an encoding that says which phenomena can count as symbols. This encoding is however as arbitrary as in the case of Shannon-information. And there is another additional problem with algorithmic information: and that is that it is always relative to a referential universal Turing machine.

Let us turn now to quantum information which is another account of information that has been successfully applied in computer science as well as in the brain sciences. Quantum mind is the idea—made famous by Roger Penrose—that considers that genuine quantum processes can (at least partly) explain the relation between mind/consciousness and the brain (cf. Penrose [22]; Beck and Eccles [23]; see Atmanspacher [24] for an overview). There have already been numerous ways proposed for how quantum processes might cause or correlate consciousness. They are all very intricate, but also controversial. The present perspective is slightly different. Can the mind/brain problem be explained with the help of an analogy to quantum processing and computers? If so, we have to ask how quantum information can be stored or encoded in a brain while being considered as a quantum system.

Quantum information is much closer to Shannon-information than to algorithmic or semantic information. The basic idea is that information is represented by quantum states which can be in a *superposition* (i.e. more states can be added – superposed together). There are many peculiarities of such states that go beyond classical mechanics, e.g. quantum states cannot be cloned (duplicated), destroyed, without the loss converted into classical states; they can be, on the other hand, teleported, i.e. transported into other underlying physical substratum. It seems, hence, that quantum mechanics allows what classical physics does not: to transfer a quantum state into another underlying medium. My response to this is that even if this were possible (and we are far from knowing it), there is no need to invoke the notion of information. There is no added value from speaking of quantum information instead of speaking of quantum states in explaining quantum processes in the brain.

In the end, we can reemploy the argument from section 2: A quantum state must be recognized or taken as a bearer of quantum information (the syntactic point (1) above). For example: quantum states A and B are realized in the spin of certain electrons. This is, however, encoding similar to the encoding of classical Shannon-information. Without such an encoding, there would be no use of invoking the concept of quantum information.

5 POSSIBLE OBJECTIONS

The Homunculus objection: The encoding agent is in the brain or in the mind. It's called a homunculus. Such a homunculus is a mind-like entity and we would need an independent explanation of its nature and skills in order to avoid a potential infinite regress. The other problem is that even if there were a homunculus that could come with the encoding, we (as external observers) wouldn't have any epistemic access to this encoding. Dennett [25] proposed that there is a way of getting rid of the homunculi regress. As complex computational operations can be analyzed or decomposed into simpler operations so the homunculus in the mind is a substantially simpler (or stupider) entity. The regress reaches its bottom when the homunculus performs simple operations. Then there can be a mechanistic account of such a way of functioning. Dennett says explicitly that this machine may be "a neurone, or a flip-flop in a computer that only has to remember 0 or 1 as its only expertise" [26, pp.87–88]. However, as we see from this brief quotation, even the machine-like homunculus operates with symbols, i.e. with bearers of information. Even if this homunculus were as simple as the neural gate discussed above, there would have to be another agent that would supply an information encoding. Converting the simplest computer into a machine thus fails. The homunculi regress is, hence, not blocked. As Searle puts it: "the only way to get the syntax intrinsic to the physics is to put a homunculus in the physics" [10, p.29].

Let us move now to another attempt of putting syntax and semantics into the physics, or better into say, in nature – to the evolutionary and teleological theories of mental content, sometimes called teleosemantics. Dretske's semantic account of information may be seen as a kind of teleosemantics. Let us first look at Ruth Millikan's influential account. Millikan claims that human brains (as well as other organs and parts of bodies) have biological purposes which are "functions that they have historically performed".

[8] The aim of compression programs is to be close to this best rate.

[27, p.330] Such functions are, in Millikan's account, the results of meme selections (there are other accounts of selections and other accounts of the origins of biological functions such as learning or conditioning).
If evolutionary design is responsible for the selection of a computational function, then the same evolutionary mechanism must be responsible for the selection of the right information encoding. If the encoding were undetermined, then it could be undetermined likewise whether the *AND-gate* or the *OR-gate* has been selected. The underlying physical states and causal structures implementing the one gate or the other are thus identical. An advocate of a teleological explanation of computational functions is, thus, committed to the claim that the evolutionary mechanism is able to select between two different features that have physical and causal powers. This means that the evolutionary mechanism could be able to distinguish non-naturalistic properties.
This initial objection against my account from a teleosemantic position can be turned into an objection against teleosemantics itself. This objection has been discussed under the label "functional indeterminacy problem", cf. Dennett [28] or Fodor [29].

Simon Blackmon recently objected to Searle's Wall argument that on the three-place notion of computation (see above), the universal realization is blocked, and the conditioning on the existence of an encoding possesses no real problem to still seeing a scientific character of the computational investigation of the mind. This argument is by an analogy to motion and location in physics and to meaning in linguistics. In physics, motion and location is assigned relative to frames of reference. In linguistics, meaning are assigned relative to an interpretation. Physics and linguistics both hold nevertheless the status of sciences. Computation is assigned relative to an encoding. Computational cognitive science holds a scientific status too. Information encoding is, thus, like a frame of reference or an interpretation. If physics is able to discover the intrinsic features of the world, then cognitive science is able to discover the intrinsic features of the brain.
This is a serious objection. I accept most of Blackmon's presuppositions: On his analysis of the notion of computation, the universal realization is, indeed, blocked and information encoding is like a frame of reference in physics or a linguistic interpretation. However, motion or location are not intrinsic features of a brain. Suppose that brain research discovers the brain centrum A is located 9 cm from the brain centrum B. This doesn't imply that the number 9 is an essential feature of the brain. We can use another unit of length and say that the distance between these centrums is 3.54 inches. By the same argument, neither voltages are intrinsic features of the brain. Arguably, different units of length don't bother physics so much, for there are simple methods of converting one unit into another one; there is no room for any indeterminacy.[9] Things get more complicated when we consider meanings and interpretations. As we know from Quine's and Davidson's well known arguments on linguistic indeterminacy, to translate one language into another or to interpret one language in another can be at times a very complicated affair. Blackmon writes:

> Significant attributions of program implementation, like those of motion and meaning, are comparative and contrastive acts. What we seek empirically is some default standard mapping, that mapping by which we can make objective discoveries of relevant similarities and differences holding not only among the brains of various biological organisms but also among artificial systems. [16]

According to this suggestion, there is a default interpretation by which we can make objective discoveries of linguistic behavior. But this idea is futile, as to interpret one's linguistic behavior is a one-time affair, cf. for example Davidson's essay "A Nice Derangement of Epitaphs" [30] for a detailed argument. The same indeterminacy concerns information encoding (except that Davidson argued for a semantic indeterminacy, while I am arguing for a syntactic indeterminacy). Information encoding is therefore more like linguistic interpretation rather than like a frame of reference in physics.

The Intentionality objection: The present argument rests on Searle's notion of intentionality. The advantage of my argument is, however, that it doesn't rest on Searle's conception of intentionality, esp. his distinction between the original intentionality, which is the feature of genuine mental states, and the derived intentionality, which can be applied onto the states of computers. This is, of course, due to the fact that I don't provide any positive account of the mind/brain relation.

6 WHAT ARE INFORMATIONAL EXPLANATIONS FOR: TWO POOL CUE BALLS

Before we proceed to the concluding section, allow me to illustrate my point that describing a system in terms of information processing is an epistemological or a pragmatic choice.
Let us have two balls on a pool table, and call them A and B. Let us assume ball A is moving and directly hits ball B. The law of the conservation of linear momentum says that the linear momentum (i.e. the product of mass and velocity) of A is transferred to the linear momentum of B (disregarding the influence of the environment). Now, assume an agent α stipulates that the linear momentum of an object encodes information (as a real number up to a certain accuracy, in gram meter per second). A ball that is not moving encodes 0, a ball weighing 10 g that is moving at a velocity of 1 m/s encodes number 10, the same ball when moving at a velocity of 0.5 m/s encodes 5 etc. Given this encoding, *we* can say that there is an information transfer from ball A to ball B after the collision. Or on colliding, ball A informs ball B about its momentum. However, an agent β could bring up another encoding by stipulating that the weight of an object encodes information. Given this encoding, *we* cannot then say that ball A informs ball B on collision. There is no information transfer, because the weights of the balls remain unchanged after a typical collision on a pool table.
On α's encoding, a collision between A and B counts as a transfer of information, but on β's encoding the same event doesn't count as a transfer of information. The crucial question is now: What description of the collision is preferable? We can describe the situation by Newton Laws (which include the law of conservation of linear momentum) or in terms of information transfer, either by α's encoding or by β's encoding (and there are infinitely many encodings). From an ontological point of view, a mechanical description is fundamental, because it is observer independent.

[9] This holds for classical (Newtonian) physics only.

There might be epistemological reasons, however, to prefer a description in terms of information transfer and furthermore practical reasons to prefer α's encoding or β's encoding.
It is convenient to describe computers in terms of information transfer, because they were designed for information processing. It might be useful to describe the brain (or its parts) also in terms of information processing. But to say that the mind/consciousness is not intrinsically biological, but rather that it is computational, is conceptually flawed.

7 CONCLUSION AND IMPLICATIONS

Let me now recap my argument: My starting point is the analogy between hardware/software and brain/mind. The brain is analogous to a computer. As computers have memory in which information can be stored, brains also must be capable of storing information. The analysis of every account of information that is relevant for this analogy shows, however, that storing information presupposes its encoding. Encoding means taking or recognizing certain physical properties (e.g. voltages in the brain) as symbols or messages. Information encoding is analogous to the syntax of a language (i.e. recognizing certain shapes as letters and words).
I argued that every encoding presupposes an encoding agent. The notion of storing information in a computer or a brain is relative to this agent (it's an observer-relative notion in Searle's words). The brain *itself* doesn't encode any information (where "itself" means "in an absolute / intrinsic / ontological / observer-independent sense"). Information encoding as well as syntax is not intrinsic to physics. Dennett proposed that what we need are "systems which store, transmit and process information in non-syntactic form" [15, p.87]. I hope to have shown that all relevant accounts of information cannot do without encoding, i.e. without syntax and that syntax cannot be accounted for in physicalist terms.
An immediate implication of this is: If the brain cannot store any information, it cannot compute any function or run any program, for non-trivial functions and programs presuppose some input information. Hence, the analogy between hardware/software and brain/mind is flawed (or it is "the most far-fetched and least useful" in Dennett's words). Nevertheless, the computational theory of the mind might be still true. The burden upon an advocate of CTM is however—following Dennett's suggestion—to provide an account of the information encoding/syntax that is observer-independent (or to get rid of the syntactic indeterminacy discussed in sect. 5). That would, in my view, involve accounting for the independent ontological reality of abstract entities (for syntax is an abstraction, i.e. recognizing particular features out of phenomena as symbols). It seems, thus, that CTM is committed to some kind of metaphysical realism in the end.
In the broad perspective, information is an epistemic notion. It is a syntactic notion (in Shannon's, the algorithmic or quantum sense) and it is a semantic notion (in Bar-Hillel–Carnap's and Dretske's sense). Using the concept of information in a non-epistemic, i.e. in an ontological sense is inconsistent with these accounts.

ACKNOWLEDGEMENTS

I'm thankful to Oron Shagrir and Marek Picha for insightful comments.

REFERENCES

[1] J. Z. Tsien. The Memory Code. *Scientific American*, July 1, 2007.
[2] A. M. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, series 2, 42: 230–265 (1936–37).
[3] J. A. Fodor. Methodological solipsism considered as a research strategy in cognitive psychology. *The Behavioral and Brain Sciences*, 3(1):63–109 (1980).
[4] D. Chalmers. The Varieties of Computation: A Reply. *The Journal of Cognitive Science*, 13:213–248 (2012).
[5] J. Searle. Is the Brain's Mind a Computer Program? *Scientific American*, 262(1): 26–31 (1990).
[6] J. Haugeland. Syntax, Semantics, Physics. In J. Preston and M. Bishop (Eds.). *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*. Oxford University Press, New York, 379–392 (2002).
[7] J. Searle. *The Rediscovery of Mind*. MIT Press, Cambridge, MA (1992).
[8] H. Putnam. *Representation and Reality*. MIT Press, Cambridge (1988).
[9] D. Chalmers. Does a Rock Implement Every Finite-State Automaton? *Synthese*, 108: 309–333 (1996).
[10] J. Searle. Is the Brain a Digital Computer? *Proceedings and Addresses of the American Philosophical Association*, 64(3):21–37 (1990).
[11] C. Shannon. A Mathematical Theory of Communication, *Bell System Technical Journal*, 27: 379–423, 623–656 (1948).
[12] F. Dretske. *Knowledge and the Flow of Information*. The MIT Press, Cambridge, MA (1981).
[13] L. Kostal, P. Lansky, J.P. Rospars. Neuronal coding and spiking randomness. *Eur. J. Neurosci*. 26(10): 2693–701 (2007).
[14] O. Shagrir. Can a brain possess two minds? *Journal of Cognitive Science*, 13:145–165 (2012).
[15] D. Dennett. *Content and Consciousness*. Routledge, London (1969).
[16] J. Blackmon. Searle's Wall. *Erkenntnis*, 78 (1):109–117 (2013).
[17] Y. Bar-Hillel and R. Carnap. Semantic Information. *The British Journal for the Philosophy of Science*, 4(14): 147–157 (1953).
[18] L. Floridi. *The Philosophy of Information*. Oxford University Press, Oxford (2011).
[19] R. J. Solomonoff. A preliminary report on a general theory of inductive inference. *Techical Report ZTB-138*, Zator (1960).
[20] A. N. Kolmogorov. Three Approaches to the Quantitative Definition of Information. *Problems Inform. Transmission*, 1(1): 1–7 (1965).
[21] G. J. Chaitin. On the length of programs for computing finite binary sequences: statistical considerations. *J. Assoc. Comput. Mach.*, 16: 145–159 (1969).
[22] R. Penrose. *The Emperor's New Mind*. Oxford University Press, Oxford (1989).
[23] F. Beck and J. Eccles. Quantum aspects of brain activity and the role of consciousness. *Proceedings of the National Academy of Sciences of the USA*, 89:11357–11361 (1992).
[24] H. Atmanspacher. Quantum Approaches to Consciousness. *The Stanford Encyclopedia of Philosophy* https://plato.stanford.edu/archives/sum2015/entries/qt-consciousness/ (2015).
[25] D. Dennett. *Brainstorms: Philosophical Essays on Mind and Psychology*. MIT Press, Cambridge, Mass (1978).
[26] A. Pyle (ed.). *Key Philosophers in Conversation: The Cogito Interviews*. Routledge, London (1999). Dan Dennett, 85–100.
[27] R. G. Millikan. Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox. *The Philosophical Review*, 99(3): 323–353 (1990).
[28] D. Dennett. Evolution, Error and Intentionality. In Y. Wilks and D. Partidge (eds.), *Sourcebook on the Foundations of Artificial Intelligence*. New Mexico University Press, New Mexico (1988).
[29] J. A. Fodor. Deconstructing Dennett's Darwin. *Mind and Language*, 11: 246–262 (1996).
[30] D. Davidson. A Nice Derangement of Epitaphs. In: E. Lepore, (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Basil Blackwell, Oxford (1986).

[31] J. Searle. Minds, Brains and Programs. *Behavioral and Brain Sciences*, 3: 417–57 (1980).
[32] J. Searle. *Minds, Brains and Science*. Harvard University Press, Cambridge, MA (1984).

Let the Machines out
Towards Artificial Social Systems

Michael Streubig [1]

Abstract: The recently revitalized branch of machine learning's has been successful in deploying bottom-up models supported by large amounts of data. Large scale simulations show new knowledge about human brain. Lehmann's theory of social systems represents a suitable foundation for the development of hybrid social systems.

1 INTRODUCTION

When Alan Turing proposed the imitation game as a method to investigate the question if machines can think, he described a social system. However, the various disciplines that have pursued this seminal enquiry rarely touch base with sociological concepts. Cybernetics developed into various interdisciplinary fields, yet it was mainly rooted in physiological models. In the meantime, the mainstream of AI focused on cognitive problem solving, predominately from a top-down approach. Traditional cognitive science rests on the concept of organisms as information processing systems – so does Artificial Life, but from a biological simulation perspective. The recently revitalised branch of machine learning has been successful in deploying bottom-up models combined with large amounts of data. Large scale simulations of the brain are expected to deliver new knowledge about the human brain. "Second-generation" cognitive science and developmental robotics are embodied and apply neural computation.

One might be tempted to say that progress has been made on brains, bodies and on models of minds. I claim that there is something largely missing in this picture, which is the social aspect. There is Social AI, and it embraces a wide variety of topics and concerns – from Stafford Beers cybernetic vision of society to simulations of interacting agents, complex systems theory, language, imitation and social learning, social network analysis and social bots, enactment, human-machine interaction, augmented and virtual environments, robot assisted therapy and behavioural game theory, to name a few. I also like to include autonomous weapons, computer worms and viruses, in particular crypto-ransomware into this context of social systems. From the other side, an interdisciplinary bridge is constructed under the label of digital sociology. The process of mutual approximation is accompanied by prolific discourses around machine ethics and emerging legal issues. A recently introduced topic of discussion is if robots should pay taxes.

So AI observes sociology, and sociology observes AI; yet they do not share a coherent theoretical program and fundamental ontological questions are still left to the philosophers. To propose an alternative route, I consider Niklas Luhmann's theory of social systems as a suitable foundation for guiding the development of hybrid social systems. A hybrid social system is understood as a social assemblage in which minds and machines mingle: humans, machines, certain things, cyborgs. Some animals are welcome, too.

To this end, I present a few selected features of Luhmann's theory and briefly visit some of their theoretical foundations: distinctions, in particular the distinction between system and environment, autopoietic systems, radical constructivism, and (second order) cybernetics. I speculate about some of the implications that arise from fostering hybrid social systems based on this particular direction of systems theory.

2 ROOTS

The following description focuses on five major influences of Luhmann's theory that are relevant for the present discussion.

Distinctions: The abstract foundation for Luhmann's theory lies in the distinction between distinction and identity. The formal background has been developed by George Spencer-Brown in his calculus of indications. While his work "Laws of Form" can be read as the description of a specific logical calculus, Spencer-Brown's intention is "proto-logical". It demonstrates that the fundamental operation of a system is drawing distinctions (as opposed to constructing identities).

System/Environment distinction: There are two fundamentally different ways to observe a system. Traditionally systems are understood by relations between their elements. In AI, these elements are constructed as agents that perceive, act and communicate. The behaviour of the system is then observed and analysed according to certain metrics, either derived from the individual behaviours or as emergent properties. The second approach, going back to Ludwig von Bertalanffy, identifies the relevant distinction for a system as the one between the system and its environment. Those systems are operationally closed and are located within the respective environments of each other.

Autopoietic systems: the term "autopoietic system" was coined by Chilean biologists Humbero Maturana and Francisco Varela. Autopoiesis denotes the capability of a system to produce and when necessary reproduce their own elements. To observe autopoiesis, the respective system references are crucial. Biological cells are autopoietic systems; they (re-)construct their

[1] Plymouth University, michael.streubig@plymouth.ac.uk

own elements. Humans grow legs, but do not re-grow a lost one. Axolotls do, however, with the help of some friends.

Radical constructivism: An epistemological position regarding the distinction between a system and its environment arises from the school of radical constructivism. According to this view, knowledge is not construed as a representation of an external reality, but as a state of the system that enables some fit with the environment. Theories of acting/enactment arrive at the similar conclusion from different premises.

Second order cybernetics: the original concept of cybernetics was developed by Norbert Wiener in his groundbreaking description of self-regulating control loops. Second order cybernetics developed these concepts further by introducing the observer into the observation. The paradigm shift is comparable to the one in physics from Newton to Einstein. It also leads to significant conceptual complication: systems of observers observing each other.

3 SOCIAL SYSTEMS

Luhmann ties up these strands into an intrinsically general theory of systems. He himself focuses on the development of a social theory based on the work of Talcott Parsons. He incorporates Spencer-Brown's concepts of distinction and indication as the fundamental operations of a system. In general, he distinguishes biological, psychic, and social systems. These systems can be structurally coupled, say, while writing a paper about systems theory. Yet spiking neurons, thoughts and an almost finished draft that is expected for publication are located in different systems that operate with fundamentally different distinctions. For Luhmann, these systems are operationally closed. They are operating in an autopoietical manner by re-constructing their own elements, and are irritated but not determined by their respective environments. This means that in Luhmann's context we cannot talk about concepts like social aspects of cognition, or observe an external environment that is part of the cognitive system. Nor can a biological system itself be social. In my opinion this separation has the benefit to be precise about the distinctions that we observe, here that the part / whole distinction has been replaced by the one between system and environment.

Social systems (interactions, organisations and institutions) and psychic systems (minds) operate with meaning. Meaning is understood as a medium in which a system can observe the distinction between actual and potential as a form. Communication can only take place through a social system. It requires the structurally coupled minds to be able to distinguish between information and utterance (Mitteilung) and to form expectations about future communication from the side of the counterpart. Luhmann also locates knowledge within the social system, rather than in the individual mind or brain.

4 HYBRID SOCIAL SYSTEMS

Let us return to the imitation game, and ask how the perspectives I have sketched out above point to hybrid social systems. I have mentioned that the imitation game constitutes a social system, one in which the participants have well-defined roles and tasks. Still it is a playful setting, which is a point that is widely overlooked. The objective for the machine is to cheat by pretending to be a woman in place of the original male participant. It is also a game that hasn't been won for the machines yet, despite the current enthusiasm for conversational interfaces.

More importantly, Turing devised the test to operationalise an idea while avoiding definitions of the concepts he had set out to investigate. Within a distinction-based approach we can be comfortable with the idea that no a priori definition is possible or needed for this kind of endeavour. Instead, we need to construct and observe the ontological theatre, as Andrew Pickering calls it.

The system/environment distinction likewise supports Turing's approach. We do not need to model a system from the biological ground up and hope it will display social behaviour. Instead, we may begin with the affordances of the social system. An agent participating in a hybrid social system should be able to act contingent on the kind of system, e.g. an interaction, organisation or institution. As an autopoietic system, it needs to be able to develop and reproduce its elements. It also requires an environment to co-develop with under evolutionary pressure. The task of the agent, its operation on a fundamental level, is to draw distinctions.

Radical constructivism suggests that social systems involving artificial agents may construct knowledge in a way that is not only structurally different from human knowledge – it may outright contradict some of our beliefs. During a debate, we treat the statements of our counterpart as opinions, not as facts. In the same manner, an artificial agent forms opinions as consequences of the autopoiesis of the system. These opinions might not necessarily be ours. In a social system, dissent does not cause logical contradictions. Instead it is processed through different selections in the medium of meaning.

Finally, artificial social agents need to be exposed to social situations in order to develop. All the machines need to do is to learn continuously without (permanent) supervision. They also need to expose a sufficiently large surface of perception and interaction: Machines that play, Robots that go hitchhiking or share our beds or roam the campus while politely avoiding humans, robots that perform art for a robotic audience, artbots that judge each other, exhibitions where humans need not apply, social bots that influence elections. These approaches are being explored and we need more of them.

When we design hybrid social systems, the key lies not in designing the system, it lies in the interaction between the system and its environment. This suggests to release the machines from the lab as soon as a minimum of functionality is implemented. The machines need to be out there and they need to be among us.

5 CONCLUSION & FUTURE WORK

Nearly three quarters of a century after Turing devised his test, the imitation game is still suitable to direct our efforts towards hybrid social systems. Even when deep learning has successfully tackled conversation (I imagine based on a similar strategy as the one for playing Go), Turing's methodology remains valid for pursuing the puzzles that remain unsolved.

Cybernetics and Artificial Intelligence have in more than one sense parted ways during the last decades. While AI has been more and more able to demonstrate practical success, it rarely reflects its epistemological foundations. Cybernetics, on the other hand, has been abandoned for the wrong reasons. In my view the complexity imposed through second order observation is a necessary condition for understanding social interactions. The question is not how to evade the problem but how to implement it. I think it will be fruitful to re-examine both paradigms, especially in the light of recent progress in robotics and machine learning. With this in mind, Luhmann's theory has the benefit to enable us to analyse, describe and construct within a coherent ontological framework that accounts for biological, psychic and social systems. Therefore I propose to model interactions between human and non-human agents as hybrid social systems. Only on the level of a social system the machine can become more human.

BIBLIOGRAPHY

[1] D. Anzola, P. Barbrook-Johnson, and J. I. Cano, 'Self-organization and social science', Computational and Mathematical Organization Theory, (Jun. 2016).

[2] D. Arnold, Ed., Traditions of systems theory: major figures and contemporary developments. New York ; London: Routledge, 2014.

[3] P. Baxter and T. Belpaeme, 'Pervasive memory: The future of long-term social HRI lies in the past', in Proc. Int. Symp. New Frontiers Human-Robot Interact. AISB, 2014.

[4] C. Beattie et al., 'DeepMind Lab', arXiv preprint arXiv:1612.03801, (2016).

[5] L. von Bertalanffy, General system theory: foundations, development, applications. New York: George Braziller, Inc, 1968.

[6] M. Bishop and J. S. Nasuto, 'Second−order cybernetics and enactive perception', Kybernetes, 34, 9/10, 1309–1320, (Oct. 2005).

[7] M. Bishop and M. M. Al-Rifaie, 'Autopoiesis in Creativity and Art', 1–6, 2016.

[8] M. Bishop, 'All Watched over by Machines of Silent Grace?', Philosophy & Technology, 24, 3, 359–362, (Sep. 2011).

[9] A. Cangelosi and M. Schlesinger, Developmental robotics: from babies to robots. Cambridge, Massachusetts: The MIT Press, 2015.

[10] A. Carlos and R. da Rocha Costa, 'On the Legal Aspects of Agent Societies', (2014).

[11] V. Charisi et al., 'Towards a child-robot symbiotic co-development: a theoretical approach', (2015).

[12] V. Charisi et al., 'Towards a child-robot symbiotic co-development: a theoretical approach', (2015).

[13] P. R. Cohen, 'If not Turing's test, then what?', AI magazine, 26, 4, 61, (2005).

[14] S. Cubitt, Simulation and social theory. London ; Thousand Oaks, Calif: Sage, 2001.

[15] A. C. da Rocha Costa, 'An Architecture for the Legal Systems of Compliance-Critical Agent Societies', presented at the Coordination, Organizations, Institutions and Norms and Norms in Multiagent Systems at the European Conference on Artificial Intelligence, Bath, 2016.

[16] K. J. Delaney, 'The robot that takes your job should pay taxes, says Bill Gates', Quartz. .

[17] M. Delvaux, 'Draft report with recommendations to the Commission on Civil Law Rules on Robotics', European Parliament Committee on Legal Affairs, Draft report 2015/2103(INL), (2016).

[18] E. A. Di Paolo, 'Overcoming Autopoiesis: An Enactive Detour on the Way from Life to Society', in Advanced Series in Management, 6, R. Magalhães and R. Sanchez, Eds. Emerald Group Publishing Limited, 43–68, 2010.

[19] B. Edmonds, 'The Inconstructibility of Artificial Intelligence by Design - the necessary social development of an agent that can pass the Turing Test', in Proceedings of the AISB'00 Symposium on Starting from Society - The Application of Social Analogies to Computational Systems, Birmingham, 33–36, 2000.

[20] R. S. Epstein, G. Roberts, and G. Beber, Parsing the Turing test: philosophical and methodological issues in the quest for the thinking computer. Dordrecht; London: Springer, 2009.

[21] H. von Foerster, 'On constructing a reality.', in Environmental design research, Stroudsburg, 2, 35–46, 1973.

[22] H. von Foerster, Ed., Cybernetics of cybernetics: "the control of control and the communication of communication" ; original edition prepared by the students enrolled in the "Cybernetics of cybernetics", a course taught by Heinz von Foerster during the fall semester 1973 through the spring semester of 1974 at the University of Illinois, Urbana, Illinois, 2. ed. Minneapolis, Minn: Future Systems, 1995.

[23] T. Fong, I. Nourbakhsh, and K. Dautenhahn, 'A survey of socially interactive robots', Robotics and Autonomous Systems, 42, 3–4, 143–166, (Mar. 2003).

[24] Z. Ghahramani, 'Unsupervised Learning', in Advanced Lectures on Machine Learning, Berlin; New York: Springer, 2004.

[25] J. W. Godwin, A. R. Pinto, and N. A. Rosenthal, 'Macrophages are required for adult salamander limb regeneration', Proceedings of the National Academy of Sciences, 110, 23, 9415–9420, (2013).

[26] M. de Graaf, S. Ben Allouch, and J. Dijk, 'Long-term evaluation of a social robot in real homes', (2014).

[27] H. Greif, 'Laws of Form and the Force of Function. Variations on the Turing Test', in Revisiting Turing and his Test: Comprehensiveness, Qualia, and the Real World, Birmingham, 60–64, 2012.

[28] F. Guerin, 'Constructivism in AI: Prospects, Progress and Challenges.', in AISB Convention, 20–27, 2008.

[29] H. Hamann et al., 'Hybrid Societies: Challenges and Perspectives in the Design of Collective Behavior in Self-organizing Systems', Frontiers in Robotics and AI, 3, (Apr. 2016).

[30] S. Harnad, 'The Turing Test is not a trick: Turing indistinguishability is a scientific criterion', ACM SIGART Bulletin, 3, 4, 9–10, (Oct. 1992).

[31] P. Hayes and K. Ford, 'Turing test considered harmful', in IJCAI (1), 972–977, 1995.

[32] P. Hingston, 'A Turing Test for Computer Game Bots', IEEE Transactions on Computational Intelligence and AI in Games, 1, 3, 169–186, (Sep. 2009).

[33] M. Hoffmann, 'Minimally cognitive robotics: body schema, forward models, and sensorimotor contingencies in a quadruped machine', in Contemporary Sensorimotor Theory, Springer, 209–233, 2014.

[34] L. Jamone et al., 'Affordances in psychology, neuroscience and robotics: a survey', IEEE Transactions on Cognitive and Developmental Systems, 1–1, (2016).

[35] R. Kibble, 'Reasoning, Representation and Social Practices', in Proceedings of the Symposium Social Aspects of Cognition and Computation, edited by R. Giovagnoli, G. Dodig-Crnkovic, and Y. Erden, 2015.

[36] B. Latour, 'On recalling ANT', in Actor Network Theory and After, J. Law and J. Hassard, Eds. Oxford: Blackwell Publishers, 15–25, 1998.

[37] B. Latour, 'A Collective of Humans and Nonhumans: Following Daedalus's Labyrinth', in Pandora's hope: essays on the reality of science studies, Cambridge, Mass: Harvard University Press, 174–215, 1999.

[38] D. Livingstone, 'Turing's test and believable AI in games', Computers in Entertainment (CIE), 4, 1, 6, (2006).

[39] N. Luhmann, 'The autopoiesis of social systems.', in Niklas Luhmann and organization studies, F. Geyer and J. van der Zouwen, Eds. London: Sage, 172–192, 1986.

[40] N. Luhmann, Social systems. Stanford, Calif: Stanford University Press, 1996.

[41] N. Luhmann, Die Wissenschaft der Gesellschaft, Nachdr. Frankfurt am Main: Suhrkamp, 2009.

[42] H. R. Maturana and F. J. Varela, Autopoiesis and cognition: the realization of the living. Dordrecht, Holland ; Boston: D. Reidel Pub. Co, 1980.

[43] M. Minsky, The society of mind. New York: Simon and Schuster, 1986.

[44] W. Myers, ‘Humans Need Not Apply Exhibition Catalogue’. Science Gallery Trinity College Dublin, 2017.
[45] P. Noriega, H. Verhagen, M. d’Inverno, and J. Padget, ‘A manifesto for conscientious design of hybrid online social systems’, presented at the Coordination, Organizations, Institutions and Norms and Norms in Multiagent Systems at the European Conference on Artificial Intelligence, Bath, 2016.
[46] K. O’Hara, ‘Trust in Social Machines: The Challenges’, in Social Computing, Social Cognition, Social Networks and Multiagent Systems Social Turn - SNAMAS 2012, Birmingham, 2012.
[47] J. Pagnutti, K. Compton, and J. Whitehead, ‘Do You Like This Art I Made You: Introducing Techne, A Creative Artbot Commune’, (2016).
[48] J. Parthemore, ‘Beyond objectification: From robots as sex toys to a new theory of personhood’.
[49] F. S. Perotto, ‘A computational constructivist model as an anticipatory learning mechanism for coupled agent–environment Systems’, Constructivist Foundations, 9, 1, 46–56, (2013).
[50] C. Pias, ‘Analog, digital, and the cybernetic illusion’, Kybernetes, 34, 3/4, 543–550, (Mar. 2005).
[51] A. Pickering, The cybernetic brain: sketches of another future, Paperback Ed. Chicago, Ill: University of Chicago Press, 2011.
[52] F. J. M. Quesada, ‘Sociology and AI: Requirements and achievements for walking towards a crossfertilization integration’, in Social Coordination: Principles, Artefacts and Theories (SOCIAL.PATH), Exeter, 58–64, 2013.
[53] S. Rainey and Y. J. Erden, ‘Turing and the Real Girl’, in The 5th AISB Symposium on Computing and Philosophy: Computing, Philosophy and the Question of Bio-Machine Hybrids, Birmingham, 56–60, 2012.
[54] J.-L. Richet, ‘Extortion on the Internet: the Rise of Crypto-Ransomware’. .
[55] M. Rocha, Cognitive, embodied or enacted? Contemporary perspectives for HCI and interaction. 2012.
[56] M. D. A. Rounsevell, D. T. Robinson, and D. Murray-Rust, ‘From actors to agents in socio-ecological systems models’, Philosophical Transactions of the Royal Society B: Biological Sciences, 367, 1586, 259–269, (Jan. 2012).
[57] A. Sadeghian, A. Alahi, and S. Savarese, ‘Tracking The Untrackable: Learning To Track Multiple Cues with Long-Term Dependencies’, arXiv:1701.01909 [cs], (Jan. 2017).
[58] A. P. Saygin, I. Cicekli, and V. Akman, ‘Turing test: 50 years later’, in The Turing Test, Springer, 23–78, 2003.
[59] M. Schiltz, ‘Space is the Place: The Laws of Form and Social Systems’, Thesis Eleven, 88, 1, 8–30, (Feb. 2007).
[60] J. R. Searle, The construction of social reality. New York: Free Press, 1995.
[61] D. Secchi and M. Neumann, Eds., Agent-Based Simulation of Organizational Behavior. Cham: Springer International Publishing, 2016.
[62] D. Silver et al., ‘Mastering the game of Go with deep neural networks and tree search’, Nature, 529, 7587, 484–489, (Jan. 2016).
[63] C. Smith, A. Matrawy, S. Chow, and B. Abdelaziz, ‘Computer worms: Architectures, evasion strategies, and detection mechanisms’, Journal of Information Assurance and Security, 4, 69–83, (2009).
[64] D. H. Smith and F. Zeller, ‘The Death and Lives of hitchBOT: The Design and Implementation of a Hitchhiking Robot’, Leonardo, (Oct. 2016).
[65] G. Spencer-Brown, Laws of form. Leipzig: Bohmeier, 2008.
[66] M. Straeubig, ‘Vom Turm in den Tümpel. Eine Annäherung an die autopoietische Maschine.’, presented at the Hyperkult, Leuphana University Lüneburg, 14-Aug-2012.
[67] M. Straeubig, ‘Can Machines Play?’, presented at the Cognitive Futures in the Humanities, Oxford, 14-Apr-2015.
[68] M. Straeubig, ‘On the distinction between distinction and division’, Technoetic Arts, 13, 3, 245–251, (Dec. 2015).
[69] A. Straschnoy et al., The New Artist. 2008.
[70] C. Teulière et al., ‘Self-calibrating smooth pursuit through active efficient coding’, Robotics and Autonomous Systems, 71, 3–12, (Sep. 2015).
[71] K. R. Thórisson, ‘A New Constructivist AI: From Manual Methods to Self-Constructive Systems’, in Theoretical Foundations of Artificial General Intelligence, P. Wang and B. Goertzel, Eds. Paris: Atlantis Press, 145–171, 2012.
[72] J. Togelius, ‘AI researchers, Video Games are your friends!’, in Computational Intelligence, Springer, 3–18, 2015.
[73] A. Turing, ‘Computing machinery and intelligence’, Mind, 433–460, (1950).
[74] M. Varga von Kibéd and R. Matzka, ‘Motive und Grundgedanken der “Gesetze der Form”’, in Kalkül der Form, 1. Aufl., D. Baecker, Ed. Frankfurt am Main: Suhrkamp, 1993.
[75] M. Villalobos, ‘Machines, Life and Cognition: a Second-Order Cybernetic Approach’, in The 5th AISB Symposium on Computing and Philosophy: Computing, Philosophy and the Question of Bio-Machine Hybrids, Birmingham, 41–47, 2012.
[76] J. Vincent, ‘Twitter taught Microsoft’s friendly AI chatbot to be a racist asshole in less than a day’, The Verge, 24-Mar-2016. [Online]. Available: http://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist. [Accessed: 11-Mar-2017].
[77] N. Wiener, Cybernetics or control and communication in the animal and the machine, 2. ed., 14. print. Cambridge, Mass: MIT Press, 2007.
[78] M. Wilson, ‘Six views of embodied cognition’, Psychonomic bulletin & review, 9, 4, 625–636, (2002).
[79] S. Woolgar, ‘Why not a sociology of machines? The case of sociology and artificial intelligence’, Sociology, 19, 4, 557–572, (1985).
[80] I. Wright, ‘The Society of Mind Requires an Economy of Mind’, in Proceedings of the AISB’00 Symposium on Starting from Society - The Application of Social Analogies to Computational Systems, Birmingham, 113–124, 2000.
[81] M. Zillich, ‘My Robot is Smarter than Your Robot: On the Need for a Total Turing Test for Robots’, in Revisiting Turing and his Test: Comprehensiveness, Qualia, and the Real World, Birmingham, 12–15, 2012.

Persons, robots and responsibility. How an electronic personality matters

Giusy Gallo[1] **and Claudia Stancati**[2]

Abstract. The recent debate raised from the motion that the Committee on Legal Affairs (Rapporteur Mady Delvaux) presented for a European Parliament resolution with recommendations to the Commission on Civil Law Rules on Robotics in order to establish rules to arrange the relationship between robots and human beings, anticipates the issue about the new kind of society in which the future generations will live in. The most relevant point of the motion is the concept of 'electronic person' to be ascribed to the robots. We will try to understand which concept of person should be used in order to define the 'electronic person', investigating the philosophical work of thinkers as Kelsen, Scarpelli, Radin, Wolff and Nèkàm. [3].

Keywords. Legal entity, legal personality, electronic person, person, robot.

1

In "Alienable rights" [1], Minsky plays out a dialogue between two aliens speaking of the rights to be granted to the human beings. He wonders if we need to allow that the world is 'pollute' with 'redundant intelligences'.

We are ever closer to the moment in which we must seriously respond to the crucial question on Artificial Intelligence: which features human beings and artificial beings share? Recently, Luciano Floridi has expressed his concern about this point: "the issue is not to decide whether the robots, as autonomous agents which are able to learn, are such a kind of person or not, but we should understand that the question is badly formulated"[2].

Actually, the question is particularly complex not only because we are considering the human being close to the artificial one, but because the term "person" is everything except that defined in univocal way. Our aim is to look for some hints in the facets of the debate on this notion in order to define the ontology of these entities which in the long run could make decisions or act autonomously.

This is the symmetrical case to the one which claims for the concepts of the extended mind and the electronic body, which have been made actual by the new technologies and have reshaped the borders of the old philosophical debate of the mind-body problem [3].

The semantic adventure of the term 'person' led the mask to what is essential: the *persona ficta* became what is essential of a subject.

The concept of the person has now entered the common language and it defines every individual in its unique peculiarity. Philosophically speaking the concept of the person has a long history which also crosses the religious tradition and particularly Christianity; today this concept has landed at the center of the discussion in bioethics, in cognitive sciences, passing through sociology, from Marcel Mauss onwards [4]. In this story, 'individual' and 'person' have been considered different concepts but only in relatively recent times they have been opposed one another: the first one has been managed such as a naturalistic category and the second one such as an ethical concept or a notion which can define a center of action.

The concept of the person, superimposed on the problem of subjectivity and strictly tied to the theme of the values and their intention acceptance, is not clear in its ethical and epistemological implications. We would mention the differences among the notion of person forged by Peter Singer or Martha Nussbaum and the role played by the ability to act, the biological nature and biography in defining those ideas.

The different notions of the person refer to different ontological levels and because of them we would like to consider here a concept of person drawn in legal field in order to verify whether this one, once opportunely adjusted, could be the right tool to define the status of robots and androids.

The crucial point of the question is to recollect a definition of the person suitable to the electronic personality with reference to the attempt of replicating human behavior, emotion and way of thinking [5], facing the issue of the imputation of liability, since the future could lead us in a dimension of co-working with robots.

2

In Roman law, the term 'persona' is equivalent to *homo*; then it evolves to an abstract meaning to designate who has a juridical role. This is a tendency which lead to the concept of the *persona ficta* or *rapraesentata* like the "fictum positum pro vero" [6, . D. 48, ii9, 16,10].

Dept. of Humanities, Univ. of Calabria, Italy. Email: **giusy.gallo@unical.it**.

Dept. of Humanities, Univ. of Calabria, Italy Email: **stancaticlaudia@libero.it**.

The authors have equally contributed to the ideas and content of this article. Claudia Stancati is responsible for sections 1 and 2. Giusy Gallo is responsible for section 3.

A very long process of doctrinal elaboration, which is parallel to the history of the Western law, particularly but non only Roman juridical tradition[4], lead to dissociate the concept of the juridical person from the legal entity, the imputation of liability from the conferring of the legal personality. The last concept does not necessarily refer anymore to the man but it's linked to the powers and the interests of the man, which are concentrated in accordance with predicative structures and normative schemes, or it refers to complex subjects to whom have been conferred features suitable to action both in the field of private law and public law.

This sort of persons are distinguished from natural elements, or rather the juridical person arises as opposed to the natural element as *persona ficta*, both with reference to a single individual and to a groups of persons. Law breaks down bodies and subjects to compose persons namely masks which could *personam sustinere o gerere, personam vicem gerere*. The same individual can be invested with different *personae*; many subjects can represent one unique *persona,* whose unity is not superimposed to the unity of a physical or psychological subject. Moreover, in some subjects there can be the rights in the absence of a dead physical or unborn person, and on the other hand the concept of physical person renames, for the purposes of the law, the biological bodies (e.g. let's consider bioethics).

The notion of legal person adapts to the definition of the artificial subjects which are allowed on a merely functional level and, then, to every artificial individual.

In his *Reine Rechtslhere* Hans Kelsen maintains that: "one might consider (…) not individuals such as, but only their actions and omissions set out in legal rules" [7] so that the physical person "is not a man but the personified unity of legal rules that give rights and duties to the same man. It is not a natural reality but a legal construction" in order to describe "legally relevant facts" [7][5].

This way Kelsen, who obviously in the time he wrote can apply law only to human beings, makes a legal person without biological constraints and he rejects the anthropomorphic view of the concept of the person.

Both the physical and the legal person are social realities such as the same law and this allow us to understand that whether "the law does not create the persons" [7][6], society can regulate a new reality described by technology using this kind of concept of person.

Uberto Scarpelli, who has introduced analytical philosophy in Italy, reached the thought of Kelsen; according to Scarpelli, the condition of use of the concept of the legal person "is in the normative forecast of a plurality of actual and possible facts, mere facts or actions, a connection among them, which consists in the reference to a same legal system and the attribution to such facts of an in operation juridical importance of relationships among them intervening"[8][7].

We can use this "limitedly technical" [9][8] concept to regulate new experiences and new realities the way it arose to answer to the new situations in which "individual subjects finds multiplied and widened their limited life in a lot of new realities" [9][9].

Also in Anglo-saxon legal literature we find similar positions. According to Radin, legal personality is "a convenient symbol, a short-hand expression by which a great many factors can be designated by a single word or a single phrase" [10, p. 625, 643] and the concept of person is "a means of making reference to complex situations easy, and is only an inexperienced person that is deceived by it, or fancies that calling the series *a*, *b*, *c*,…*n*, by the name of x, add a new item to the series"[10, 665]. Wolff speaks about abbreviation and a formula of fiction. Starting from the natural extra-legal concept of personality and of entities considered such as persons despite the lack of human dignity[11, 494, 505].

Rules are a fundamental mechanism of connection between the social and the mental by the means of the so-called "multiagent structure" that expresses itself in the plurality of the recipient of the majority of legal rules. This structure allows an education which falls beyond the one imagined by Kant, for whom from the concept of legal entity emerges only and necessarily the relationship between free and equal people and for this reason, according to him, legality cannot connote neither our relationship with animals nor the one with God or, we can say today, with the androids. However, the process of de-psychologizing, abstraction, de-naturalization that features the rule of laws allows to them to refer to different models of social organization and it shows how they are semantically multipotent thanks to the ability to use natural elements in an artificial context, in which sometimes by the means of the rule "*x* counts as *y* in the context *c"*[12] it occurs the transition from the physical object to the social one.

According to Nèkàm the feature that makes a legal entity is the fact that the collectivity considers it as a unity which has interests that need or deserve social protection: "and it is completely irrelevant that such entity exist as an objective reality, that has its will, and that it possess or less a natural or artificial, imagined or proven personality"[13, 21]. Following this perspective, we can think to look for the theoretical grounds to reshape the new legal challenges that Artificial Intelligence begins to set.

---

[4] The term 'person' is not present in the *Code Napoléon*, but one can find it in the *common law*, as it shows the *Judicial Dictionary* by Frederick Stroud [18].

[5] Translation from the Italian: *La dottrina pura del diritto*, Torino, Einaudi, 1952, p. 193.

[6] Translation from the Italian: *La dottrina pura del diritto*, Torino, Einaudi, 1952, p. 198.

[7] Translation from the Italian: *Contributo alla semantica del linguaggio normativo*, Milano, Giuffrè, 1959, p. 142.

[8] Translation from the Italian: *Il problema della scienza del diritto*, n. ed. ac. P. Piovani, Milano, Giuffrè, 1962, p. 81.

[9] Translation from the Italian: *Il problema della scienza del diritto*, n. ed. ac. P. Piovani, Milano, Giuffrè, 1962, p. 82.

The recent developments in the field of Artificial Intelligence, particularly in robotics, claim for a philosophical approach to the newborn (and future) relationships between human beings and robots. The continuous online life we are experiencing is shaping our way of living in the information society asking – as Luciano Floridi [14] underlines – for a new ethical framework, in which we should ward off the anthropocentric view over robots allowing to accept them as co-workers, including the huge amount of legal, social and ethical issues at the stake.
Despite the fight between cyber-optimists and cyber-pessimists [15], it is no longer possible to image our life without Artificial Intelligence: algorithms follow and support us in a series of daily actions (withdraw from an ATM, watching a TV series on Netflix, booking a doctor's appointment, applying online at university and so on). Also if we accept these kinds of examples of Artificial Intelligence, generally speaking, we are skeptics about robots and androids. Because of science fictions films and TV series, one might conceive human world invaded by robots, condition which presupposes a fight for supremacy won by artificial systems created by the losers human beings. Obviously this is a naïve and false view on Artificial Intelligence, since we are not facing an invasion, but those artistic perspective could give us some hints in order to imagine the new challenges for the future infosphere [16].
Science fiction showed us the chance to have driverless cars (or any other transport). Now, the project of a driverless car is under construction: Google, Tesla, Uber and Nissan are testing their prototypes of driverless or autonomous car. Although there occurred little accidents, there are recent news about the capability of autonomous car to prevent bigger accidents, like the one prevented in December by a Tesla car running for testing along a Dutch highway[10]. Let's consider for a while that in the same situation, there occurred a malfunction of the 'Forward Collision Warning' onboard the Tesla car: instead of preventing the accident, the autonomous car provokes it, investing a car carrying three people, including a child. The result of the accident is little damages to the three people and huge damage to their car. Who is responsible for this accident? This is the real question at the point of the robotics revolution. There is no need to be afraid of robots, since they could help us with machine learning, overall in medicine for application in health systems for improvement in cancer and other rare diseases research.
The proposed example shows that we need to think the relationship we would like to experience with those artificial systems. Following this perspective, the Committee on Legal Affairs (Rapporteur Mady Delvaux) presented a motion for a European Parliament resolution with recommendations to the Commission on Civil Law Rules on Robotics[11] in order to begin to establish rules and a kind of ethical and legal code to arrange relationships between human beings and Artificial Intelligence artifacts. The main issue is to protect persons from the damages eventually provoked by Artificial Intelligence systems. Stressing the increasing production and use of Artificial Intelligence systems, the motion faces legal, ethical and social issues. Luciano Floridi [14] offers us a clear analysis of the motion, stating that its most relevant features are the issue of unemployment and liability. We could add to these issues that developments in robotics and artificial intelligence can and should be designed in such a way that they preserve the dignity, autonomy and self-determination of the individual, especially in the fields of human care and companionship, and in the context of medical appliances, 'repairing' or enhancing human beings [17, p.5].

This idea covers the chance that developments in robotics and AI could be helpful to people but their application need to be constantly checked since one cannot ignore the users are not simply individual but persons which sometimes have a reduced autonomy but they have the right to preserve their dignity.
Back to the two features stressed by Floridi, we should recall the example of the driverless or autonomous car. Let's say, for example, that within a couple of years we could travel with a Google or Tesla car. It makes sense that taxy service companies might fire their drivers, just like bus and train companies since there is no more need to pay people for a job they don't do anymore. So that there might be a lot of unemployers. If this scenario should happen, this means that the social relapse will be dramatic, so that today there is already the need to think about an economic support for people who will lose their job because of robots and algorithms. Then, companies will save money and will reduce complaints, but the social systems of some Nations will be troubled by this new developing situation. The only solution is to tax robots, which means to tax the companies which use them and replace a lot of human workers with robots.
What if an autonomous car crashes against another car drawn by a human being? Who is legally liable for the accident? Apparently nobody or, at least, the company who has produced the car if there is a malfunction or the owner who made a mistake in running the system. However, the main question concerns whether our legal systems are ready to face these situations and what are the consequences on the notion of the person.
The previous paragraph showed that there is not a univocal notion of the person. Since concepts such as rights, responsibility, ethics, privacy have been ascribed to persons and not to material things such as robots, how to shape the concept of the person?
The motion proposed to the European Parliament seems to solve the problem creating the notion of 'electronic person', a new legal status to be given to autonomous robots:

creating a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently [17, p. 18].

Unfortunately, the term 'electronic person' appears only once in the whole motion (the last version is a document of 68 pages). This recommendation to the European Parliament is not deeply explained or investigated.
It seems that the legal commission, taking into account the difficulties in investigating the notion of the person as it is clear in the history of the philosophy of law, has proposed the notion

[10] Link: http://www.carscoops.com/2016/12/teslas-autopilot-anticipates-huge-crash.html.

[11] Link: http://www.europarl.europa.eu/committees/en/juri/subject-files.html?id=20170202CDT01121.

of ‘electronic person’ in order to begin to think to the consequences of robots on human beings.

The purpose of the commission which uses the term ‘electronic person’ seems a mistake: since an ‘electronic person’ is a robot, then it is an artifact, which is sometimes, in particular, an Information Technology artifact, and it has to be considered an artificial system. Unfortunately, this artificial system could be invested of rights, duties, legal liability. If this should happen, then it will make sense to start thinking about robots such as Isaas Asimov’s *Bicentennial Man*, ascribing to robots – in future – also emotions, according to the most wide meaning this word can cover.

The anthropomorphic view risks to make us think the problem of the relationship between robots and persons – the persons we are – following the path which sometimes we run thinking about the relationship between human beings and animals. However, there is a big question: animals are not artificial systems, but robots behavior is always governed by rules imposed by their creators.

We can say, agreeing with Luciano Floridi, that the relevance of the issue needs to be considered changing the point of view. So then, we need to reshape the notion of electronic person, giving sense to the terms electronic and person in a philosophical perspective. In particular, in the previous paragraphs, we showed that the philosophy of law of Scarpelli, Kelsen and Nèkàm, for example, ward off the biological constrain which generally is ascribed to the person, trying to describe a new kind of society in which there is a widespread co-working with robots and androids.

REFERENCES

[1] M. Minsky. Alienable rights. In *Android epistemology*, Cambridge MA, MIT Press (1995).

[2] L. Floridi. L’ultima legge della robotica. In: La Repubblica, published 12.2.2017.

[3] G.Gallo and C. Stancati. Where is the mind? The extended mind reloaded. In: AISB Conference Proceedings, Sheffield 2016.

[4] M. Carruthers et al. (eds.). *The Category of the Person. Anthropology, Philosophy, History*, Cambridge, Cambridge University Press (1985).

[5] M. Minsky. *The emotion machine. Commonsense thinking, artificial intelligence and the future of the human mind*, New York, Simon&Schuster, Cambridge MA, MIT Press (2006).

[6] Bartolo di Sassoferrato, *Commentarium in Digestum Novum*, c. D. 48, ii9, 16,10.

[7] H. Kelsen. *Reine Rechtslhere*, *Reine Rechtslehre*, Wien, F. Deuticke (1934).

[8] U. Scarpelli. *Contributo alla semantica del linguaggio normativo*, Milano, Giuffrè (1959),

[9] G. Capograssi. *Il problema della scienza del diritto*, n. ed. ac. P. Piovani, Milano, Giuffrè (1962).

[10] M. Radin. The Endless Problem of Corporate Personality. *Columbia Law Review*, 32 (1932).

[11] M. Wolff. On the Nature of Legal Persons. *The Law Quarterly Review*, 54 (1938).

[12] J. Searle. *Mind, Language and Society*. New York. Basic Books (1998).

[13] A. Nékàm. *The Personality Conception of the Legal Entity*, Harward Studies in the Conflict of Laws, Cambridge, 1938.

[14] L. Floridi. Roman law offers a better guide to robot right than sci-fi. Financial Times. 22.02.2017.

[15] G. Lovink. (2012). *Network without a cause*, Polity Press (2012).

[16] L. Floridi. Robots, jobs, taxes and responsibilities. *Philosophy and Technology*. DOI 10.1007/s13347-017-0257-3 (2017).

[17] Committee on Legal Affairs (Rapporteur Mady Delvaux). European Parliament resolution with recommendations to the Commission on Civil Law Rules on Robotics: http://www.europarl.europa.eu/committees/en/juri/subject-files.html?id=20170202CDT01121.

[18] F. Stroud. *The Judicial Dictionary of Words and Phrases Judicially Interpreted*. London, Sweet & Maxwell, Limited (1890).

Autonomy and Social Attitudes

Raffaela Giovagnoli [1], Josefina Formanova [2]

Abstract. I'll follow a line of thought that suggests to intend "personal autonomy" in a social sense. The urgency to undertake this move arises because of the wide variety of informational sources we are exposed which influence our behavior. Social background represents the basis for autonomy; at the same time, interaction with others (real or virtual) enlarges the possibility for autonomous judgements. My attempt is to try to elucidate the connection between autonomy, knowledge by testimony and the exposition to informational diversity.

1 INTRODUCTION

The notion of "personal autonomy" defines those philosophical views that prefer to intend autonomy not only in a "moral sense". Moreover, my discussion does not concern the classical debate on free will (determinism/indeterminism). There are several theories of personal autonomy, that suggest to overcome the "individualistic" views (Giovagnoli 2007). Nevertheless, they are "social" in different senses, which I'll briefly describe.

Some philosophers maintain that an autonomous person must show not only "procedural" but also "substantial" independence, which rightly requires the consideration of the social context in which an action can be judged as autonomous. Procedural personal autonomy requires the fulfillment of conditions for rationally deciding and acting, ranging over a wide spectrum of individual idiosyncratic desires and volitions (Dworkin, Frankfurt, Ekstrom). Several authors try to add "historical" conditions which examine the process of the formation of judgment or decision to act (Fisher, Ravizza, Christman). Very interesting and powerful are the substantive theories, that in the individual variant start from the Kantian inheritance (Korsgaard, Hills, Wolf).

My argument focuses on substantive views which consider the social contexts as sources of norms which favor or diminish development and exercise of the capacity for personal autonomy.

2 SOCIAL CONCEPTIONS OF AUTONOMY

The Habermasian account of "communicative action" represents a good model to relate autonomy to a linguistic normative competence. According to Habermas, autonomy is bound to the acknowledgment of presuppositions or linguistic rules as conditions of universal validity of theoretical and practical claims. Differently from other theories of "recognition" (nota), he plausibly introduced the issue of "interpersonal recognition" related to formal linguistic conditions of a rational and egalitarian dialog [7]. Habermas account of the relationship between autonomy and socialization is convincing because it shows plausible arguments against the contemporary reductionist strategies. Autonomy is directly bound to the notion of "communicative action", which, differently from instrumental and strategic action, aims at a rational consent. It is not only a question of an ideal point of view under procedural conditions (introduced from an external perspective), rather the development of the possibility for autonomy through communicative action is based on a well-known process of socialization explained also by reference to different disciplines (Piaget, Kohlberg).

To become an autonomous individual being, i.e. to become a self, means for Hegel in Brandom's interpretation (Brandom 1999) to take it or treat it as one by those that are taken or treated as ones by this very being. For the individual autonomy depends on the condition of being recognized as one simultaneously with recognizing the other (treating me as one) as one as well, the process may be called a *mutual recognition* (based on Hegel's *phenomenological viewpoint*). Under such a treatment the recognized one is fully ascribed the status of capability of being *responsible* and of *becoming a subject of commitments*. To reach such a status is the kind of *social achievement* that conditions the individual's autonomy. This would, however, sound too restrictive if we admit that one's autonomy depends strictly on the others' decision to ascribe her this kind of freedom. Therefore, there is one more condition that makes this definition sound less paradoxically constrained. Speaking of autonomy one shall also take to account the fact that the one's ability to take responsibility also means the one is entirely free to reject it. "According to the autonomy thesis, one is in a strict sense bound only by rules or laws one has laid down for oneself, norms one has oneself endorsed." (Brandom 1999)

Procedural theories [5] underestimate the role of the internalization of oppressive norms; from a substantive point of view, even if a person has to some extent the option of choosing alternative values, it is the content of the norms she internalized that diminishes her autonomy. We must consider the nature of values that come from the social context (Stolyar, Benson, Oshana among others). For instance, Benson focuses on a fundamental point in order to understand the normative source of autonomy. He addresses directly to the social and discursive dimension of "taking ownership" that explains how internalized invisibility (internalization of oppressive norms diminishing autonomy) can defeat agents' capacity "to take ownership" of what they do [6]. The active dimension of taking ownership implies the capacity of the agent of giving reasons for her actions and so of responding to potential "challenges" arising in the social context from her own point of view. This move gives the agent the possibility of taking responsibility in a public context and also the possibility of speaking for people who are marginalized.

[1] Faculty of Philosophy, Pontifical Lateran University, giovagnoli@pul.it

[2] Faculty of Philosophy, University of Prague, fina.formanova@gmail.com

3 AUTONOMY AND COMMUNITY

Several relevant authors in the field of social epistemology think that the model for an ideal knower must be abandoned (see Hardwig, Welbourne, Schmitt, Baier, Webb, Goldman, Jones, Fricker Faulkner, Lipton, Kusch, Lackey). These authors generally think that we cannot neglect ethical and social dimensions of inquiry. A strong view among social epistemologists is that an individual cannot be said to know, via testimony, that p, unless p is known in the community (see Welbourne, Brandom, Faulkner). The autonomy of individual reason become a chimere.

In Brandom's account of "recognition" in scorekeeping terms the possibility to undertake an autonomous perspective is bound to "deontic attitudes" and "deontic statuses". Recognition seems a fundamental requirement for an agent to be autonomous, but reciprocal recognition is possible by virtue of "shared commitments". Beside the basic form of "simple recognition" through which we recognize each other by virtue of our common intentionality toward natural environment, self-consciousness requires "robust" recognition. Simple recognition entails to have a conception of the self in a double sense. First, an agent recognizes herself as something as a self, i.e. as able differentially to respond to environmental stimuli moving from the satisfaction of basic desires (such as the attitude of "hunger"). Second, an agent must have a capacity for *recognition*. She must have a conception of the self as able to do what is required in order to be taking or treating something as a *self*, namely a subject of normative statuses of *authoritative* (in the sense of probative, though still provisional and defensible) *commitments* as to how things are. This second sense implies reflexivity, because the agent recognizes oneself among those who she recognizes. Starting from simple recognition, the agent can reach the dimension of "robust" recognition if she is disposed to acknowledge the simple recognitions of others and undertaking the consequent commitments by herself. In Brandom's terms: «Thus robust self-consciousness is achievable only through *reciprocal* recognition: being robustly recognized by at least some of those one robustly recognizes. This means that a *community* (a kind of universal) is implicitly constituted by one's own robust recognitions, and actually achieved insofar as they are reciprocated. That is the sort of reciprocally recognitive community within which alone genuine (robust) self-consciousness is possible: the "I" that is "We" and "We" that is "I"»[3].

The result is that autonomy is not thinkable without shared knowledge and testimony is a fundamental source of knowledge. An interesting observation simply suggests that an epistemic agent whose beliefs does not depend upon testimonial transmission knows very little. So, we must consider our limitations and our need to economize. If we do not ignore these, autonomy will appear irrational for accepting a huge loss of information, ignoring how we exploit a division of epistemic labor (Putnam, Kitcher).

4 HOW TO AVOID THE PARADOXICAL CONSTRAINT OF AUTONOMY

As was remarked, Brandom assesses on the so-called *autonomy thesis* developed by Kant based on his Rousseau reading (Brandom 1999). On the one hand, there is the *authority* of the others' over me that is, on the other hand, constrained by my acknowledgement of this authority. At this stage, the autonomy thesis seems symmetrical. Autonomy thesis means basically that one has the capacity to bound herself by norms, i.e. to *apply* certain *concept* according to the norm. But this capacity of mine depends on the others to lay the responsibility towards to norms, according to which I apply the concept, upon me. This, in Brandom's view merely *seemingly* paradoxical, constraint may, however, cause huge difficulties once we want to hold the idea of autonomy in a strict sense of the term. My authority over my activity seems real but partial. Is it possible to keep the idea of autonomy sensible despite its partiality? How are we, therefore, to solve this paradox?

Let us sketch an example that should enlighten the paradox in terms of Wittgensteinian conception of *language game*. If one wants to join certain language game, i.e. respond to an imperative: "Do the dishes!" correctly, what shall the one get to know and embrace. First, one need to understand the sentence, i.e. understand the language both structurally and semantically. That alone requires the one to has apprehended certain social background already. Second, one need to decide whether to listen to the imperative and yield to it. This analysis shows that even though the one rejects to obey the order or request, there remains one condition, i.e. the first concerning her understanding of the order, that disturbs the balance of the whole recognitive structure. The second condition represents the exact symmetrical structure Brandom offers in his reading of Hegel: one appreciates certain language game, norm respectively, but lays his authority over it by making a free decision to apprehend it at the same time. Yet, the first condition, as was concluded, does not possess this symmetry, and therefore makes the whole idea of autonomy fallacious.

In *On Certainty* Wittgenstein clearly states that there is no such thing as the last belief that would ground all the other beliefs so to say. Similarly, if we infer from this argument, we could say there is no such thing as the last language game either. Davidson states there is a kind of interdependency between all beliefs. The *conceptual*, including the capacity of *naming* objects and *forming* beliefs needs to rely on some other ground, therefore. This ground is the *social* ground, for if it were merely *empirical*, the argument would result either on *solipsism*, or on the idea of the last grounding belief which has been rejected already. On this behalf, the paradoxical state of autonomy thesis should be solved as well. Language games, i.e. norms in our terminology, starts from certain social interaction. But, every social interaction always presupposes certain language or norms pre-established already. To speak about the beginning of language or sociality is, therefore, misleading, and only resulting in an *infinite regress* (Formanova 2015). To erase the problem of asymmetry of autonomy thesis entirely, we shall, however, explain, how the language and the sociality works in time.

Negotiating and adjudicating the claims of mutually conditioning authorities, administering conceptual norms by applying them in distinct cases etc., is a *process* (Brandom 1999). Applying and instituting (administering) conceptual norms is not merely a temporal process but a *historical* one. That is what Hegel enlighten in his *Phenomenology of Spirit*, and what results from the fact that all that happens in time is socialized. The future application of certain concept has an authority over the past ones by assessing them in terms of their

---

[3] Ivi, p. 32.

correctness or incorrectness. Simultaneously, the past applications have their authority over the future ones while holding the tradition according to which the future applications is be made, or at least respond to it in some way. Therefore, the present application, as a mediator between the future and the past ones, responds and lays its authority twice. Once towards the past applications of the concept with accord to or critical assessment of the past ones; once towards the future applications that are pre-established both in the past and the present ones, and which are at the same time authoritative towards the previous two by featuring as their future assessor and corrector. In these relations, we may see how the mutual recognition works in general. For anytime X has authority over Y, X needs Y in order to have something that would hold its authority on it. Therefore, both X and Y are equally important in this relation.

5 CONCLUSION

Recognition is recognition of inferentially structured commitments[4]. This thesis implies that I am not forced to follow commitments that are valid for a person or a group of persons or a community. If a person was educated in an open and flexible manner then she has concrete chances to be autonomous because she becomes ready to take part to the game of giving and asking for reasons, and this implies the consideration of all the reasons I can plausibly come to know.

REFERENCES

[1] Benson P. (1991), Autonomy and Oppressive Socialization. *Social Theory and Practice*, 17: 389.
[2] Benson P (1994), Free Agency and Self-Worth. *Journal of Philosophy*, 91, 650-668.
[3] Brandom R. (1994), *Making It Explicit.* Cambridge University Press, Cambridge.
— (2007) The Structure of Desire and Recognition. Self-consciousness and Self-constitution. *Philosophy and Social Criticism*, 33: 127-150.
[4] Christman J. & Anderson J. (Eds.) (2005), *Autonomy and the Challenges to Liberalism*. Cambridge University Press, Cambridge.
[5] Christman J. (1991), Autonomy and Personal History. *Canadian Journal of Philosophy*. 21: 1-24.
[5] Fisher M.J. & Ravizza S. J. (1998), *Responsibility and Control*. Cambridge University Press, Cambridge.
[6] Giovagnoli R. (2003), On Normative Pragmatics. *Teorema*, XXIII, 51-68.
[7] Giovagnoli R. (2007), *Autonomy. A Matter of Content*. Firenze University Press, Florence.
[8] Habermas J. (1990), *Post-metaphysical Thinking*. The MIT Press, Cambridge.
— (1998), *The Inclusion of the Other*. The MIT Press, Cambridge.
— (2008), *Between Naturalism and Religion*, Polity Press, Cambridge.
— (2007), "The Language Game of responsible Agency and the problem of Free Will: How Can Epistemic Dualism Be Reconciled With Ontological Monism?" *Philosophical Explorations*, 10: 57.
[9] MacEnzie C. & Stolyar N. (Eds.) (2000), *Relational Autonomy*. Oxford University Press, Oxford.
[10] Taylor J.S.(Ed.) (2006). *Personal Autonomy*. Cambridge University Press, New York.
[11] Oshana M. (2006), *Personal Autonomy in Society*. Ashgate, Hampshire.
[12] Owens D. (2000), *Reason Without Freedom: The Problem of Epistemic Normativity,* Routldge. London.

[4] This theoretical opinion differs from the Hegelian interpretation of recognition and the space of reasons offered by Italo Testa in his essay *Seconda natura e riconoscimento*, in *Lo spazio sociale delle ragioni: da Hegel in avanti, Guerini, Milano, 2007 forthcoming.*

Artificial Life, Artificial Unconscious

Mohammad Ali Rahebi

Abstract

The Unconscious is the Body

Michel Serres

Cybernetics and the Cartesian Problem

The way the body works without knowledge or consciousness has been a complex and certainly consequential problem since Descartes at the latest. The question of how we do something without knowing *how* we are in fact doing so was a scandal that drove Cartesians to Occasionalism (a concept that is perhaps more true now than ever, with cloud computing and distributed computation).

Since then, we have come very far indeed but it is in AI that we come to again feel the real force of this problem again. Minsky has famously stated that

Social cognition *qua* language based and mainly representational, is a highly inefficient mode of *signal processing* or computation in general. As Minsky (despite his own strict adherence to the representational models, strangely enough) said:

> *It's mainly when our other systems start to fail that we engage the special agencies involved with what we call "consciousness."*

As I shall try to while discussing the Peircean notion of "community of believers" and the way the Cybernetic schema manages to overcome its necessity, the most efficient manifestations of artificial intelligence are not representational, as is the case with the recent success of neural networks and machine learning. In fact, if we are to investigate artificial *life* instead of artificial *consciousness* (a failed project; consciousness *qua* delay is simply obsolete in machinic terms), we have to look at the body, at the *habitual*, that is to say at the non-representational, non-knowledge-based mechanisms and automations that are not language-based in the least.

Beyond Turing

In a relatively recent book edited by Raffaela Giovagnoli, one of the organizers, *Computing Nature,* the issue of alternative computation models to Turing Machines was broached and discussed to some extent. Here I will argue that Cybernetics is one such alternative model that has been growing, from its origin as a strange interdisciplinary field in the late 1940s, to become the dominant model of computation. Although it is not named as such, the models that are not algorithmic and termination-oriented are all operating on the basis of the "cybernetic schema of intelligibility." These are very familiar processes for all of us, being ubiquitous in "smart" machines and smart software. Giovagnoli et al describe Cybernetics as computational processes where

> *The main criterion of success of this computation is not its termination, but its behavior - response to changes, its speed, generality and flexibility, adaptability, and tolerance to noise, error, faults, and damage.*

If consciousness is to be taken as linguistic and representational, and as such mediated by the social in its historicity and theoretical bias, Cybernetics would have to be seen as an artificial unconsciousness, and in fact, is much closer to artificial *life* than such AI trends as GOFAI or even most strands of embodied cognition.

In French philosophy, under the influence of Derrida and his reading of Plato, the problem of technology has been tied to that of writing, as some form of Ur-Technics that contains the essence of all future technics *qua* mnemotechnics. This idea has been also taken up by Bernard Stiegler who has taken it further by discussing all technologies as means of "retention" and "protention." All of this stars with Plato, however, for in his famous Phaedrus, he comes to barrage writing as a dangerous method, a supplement to spoken language, which unbeknownst to its to its users, had deleterious effects on memory as well as on truth and on the *polis*. We are not here concerned with all that; what is of interest is something he mentions in passing, namely that if you were to write down something that is true at one moment ('it is day'), it would become false the next (as night fell) because unlike the speaker, it did not the ability to correct itself according to its surrounding reality, its environment. In the same manner, he says that when you write down the teachings of some philosopher, the written text will not be able to answer new questions or clarify obscure passage when asked to do so, in comparison to the person using living language.

This is in fact the same thing that distinguishes Cybernetics completely from all the technics that came before: cybernetic machines (think of your smartphone) *can* change their displayed content *and* their behavior in response to changes in their environment. Unlike previous forms of technology, which might have been mnemotechnics (or not), cybernetic technologies are adaptive, self-modifying, robust. In fact, this same Platonic problem occurs in Descartes (of whom we will not have time to speak) and later, Turing. In his famous "On Computation," Turing states that although certain numbers might be

uncomputable for the Turing Machine, they might be very well computable for human mathematicians. The reason for this is that the human mathematician is capable of revising and changing "their strategy" completely from the ground up, while the Turing Machine, the representational computing machine, cannot change its own behavior when faced with an unsolvable problem. It does not have the ability to *change* its actions *spontaneously* and in response to its specific problem-environment. This is of course recognizable as the same Platonic problem that we encountered with writing and other technological artifacts. They are Leibnizian in principle, relying on some pre-established harmony, on the stability of the operations in the environment and thus are rendered absolutely inoperative once the smallest change occurs.

The cybernetic machine, moreover, does not start from a human-defined state or representation, but is an immanent, non-representational computing machine whose operations (e.g. in case of a neural network running Big Data analyses) cannot be even comprehended by a human observer. As long as there be no need for the cybernetic machine to interact with a human, there is no need for representational information; data is much more efficient.

Peirce, Habits, Neural Networks

Finally, I will present some of my current research on the relation between the American Pragmatist philosophies, the very important yet under-valued concept of *habits*, and the recently most successful manifestation of the cybernetic schema, namely Neural Networks.

The importance of the Pragmatists lies in their discovery of the importance of habits (for they too were early fans of feedback mechanisms, having become familiar with the pre-cybernetic feedback control called the flywheel governor). Yet, though their philosophy managed to move to pure empiricism, it could not solve the notion of intelligent encounter without recourse to social mediation and meaning-creation. This is now solved by the creation of neural networks that are in fact close analogs of Peirce's community of believers if considered as a community of nodes holding specific weights and bias parameters.